METHEXIS

La scienza e la rete

L'uso pubblico della ragione nell'età del Web

FRANCESCA DI DONATO

Francesca Di Donato

La scienza e la rete

L'uso pubblico della ragione nell'età del Web

Firenze University Press
2009

A Giulia, Michele e Martina

SOMMARIO

INTRODUZIONE 9

CAPITOLO 1
IL SISTEMA TRADIZIONALE DI COMUNICAZIONE NELLA SCIENZA 13

1. La nascita della comunicazione scientifica moderna 13
2. I sistemi di comunicazione della scienza e la valutazione del sapere 20
3. La crisi del sistema tradizionale e le trasformazioni in atto 29

CAPITOLO 2
UN SISTEMA UNIVERSALE DI COMUNICAZIONE:
IL WORLD WIDE WEB 33

1. Un po' di storia 33
2. Dal Mundaneum al Memex: alcune premesse teoriche 43
3. Architettura del World Wide Web e architettura dell'informazione 51

CAPITOLO 3
WEB DEI DATI E SOCIAL SOFTWARE 59

1. Cenni di topologia delle reti 59
2. Oltre i limiti del Web: il Web semantico 66
3. Ragnatele sociali: dal Web 2.0 al Web 3.0 80

CAPITOLO 4
FARE SCIENZA IN RETE 91

1. Un ritorno all'antico: dalle RFC al software libero 91
2. L'avvento del movimento Open Access 101
3. Fare ricerca sul Web 3.0 108

CONCLUSIONI 121

INDICE DEI NOMI 125

BIBLIOGRAFIA 129

Francesca Di Donato, *La scienza e la rete : l'uso pubblico della ragione nell'età del Web*,
ISBN 978-88-8453-494-1 (print) ISBN 978-88-8453-500-9 (online)

INTRODUZIONE

L'avvento e la diffusione su scala planetaria delle reti telematiche (Internet e il World Wide Web) hanno radicalmente trasformato le modalità di comunicazione. La rete, intesa sia come infrastruttura tecnologica sia come comunità sociale globale, è lo spazio in cui, in misura sempre crescente, si creano e si disseminano informazioni, dati, conoscenze e competenze.

Questo libro considera un ambito o "campo" particolare della comunicazione in rete: quello relativo alla scienza.

Nonostante questa preliminare restrizione di campo, il tema qui affrontato non riguarda una questione specialistica o settoriale, che interessa solo il mondo accademico. La scienza infatti è un *commons*, un bene di tutti, che esce dall'accademia attraverso le aule delle università, e tramite la divulgazione dei suoi risultati raggiunge, oltre ai ricercatori stessi, diversi destinatari: imprese, insegnanti, politici e cittadini. La scienza, si può dire, riguarda tutti.

Affronteremo il rapporto tra la scienza e la rete a partire da due questioni preliminari.

Primo: perché la società ha bisogno di una scienza libera? A questo, tra l'altro, risponde Kant nella *Risposta alla domanda: che cos'è l'illuminismo?* del 1784[1], alla vigilia della Rivoluzione Francese. L'uso pubblico della ragione, l'uso cioè che uno fa del *logos* «*in quanto studioso*, davanti all'intero pubblico dei *lettori*» [A 485], dev'essere libero, perché solo così può attecchire l'abitudine al pensare da sé, al *Selbstdenken*. Esercitare l'uso pubblico della ragione, se è difficile per il singolo individuo, è possibile per il pubblico. Il rischiaramento può avvenire collettivamente, come fenomeno storico-culturale, una volta che al singolo sia lasciata la libertà di ragionare pubblicamente su scienza, religione e politica; anzi, aggiunge Kant, che ciò accada è «persino inevitabile». Affinché i singoli possano pensare autonomamente, dev'essere possibile discutere pubblicamente di scienza. Solo così può germogliare la vocazione al libero pensiero, il quale «allora agisce a sua volta gradualmente sul modo di sentire del popolo (attraverso la qual cosa questo diventerà più e più capace della *libertà di agire*), e alla fine addirittura sui princìpi del *governo*» [A 493-4]. La ri-

[1] I. Kant, *Risposta alla domanda: Che cos'è l'illuminismo?* (tr. it. di F. Di Donato), «Bollettino telematico di filosofia politica», <http://bfp.sp.unipi.it/classici/illu.html> (ed. orig. 1784).

Francesca Di Donato, *La scienza e la rete : l'uso pubblico della ragione nell'età del Web*,
ISBN 978-88-8453-494-1 (print) ISBN 978-88-8453-500-9 (online)

sposta di Kant alla nostra prima questione è che la libertà della scienza è condizione della libertà politica. In termini più attuali, potremmo dire che la libertà della ricerca è un prerequisito di una società liberale e democratica.

La seconda questione riguarda il rapporto tra lo statuto della scienza in una data cultura e il modo in cui è prodotta e veicolata. È un tema legato alla stessa nascita del pensiero filosofico occidentale, sorto in un'epoca di rivoluzione mediatica che ha visto il passaggio dall'oralità alla scrittura. Platone lo affronta nel *Fedro*[2], un dialogo che ha al centro la comunicazione scientifica, affermando che i contesti in cui si determina il modo in cui si usano le tecnologie di comunicazione condizionano i gradi di libertà della scienza. Platone dunque afferma che il mezzo non è neutro, presenta limiti e possibilità. Ma l'effetto che produce dipende da come lo si usa, all'interno di questi limiti.

Nel *Fedro,* in risposta al Mito di Teuth (274e-275a), Socrate afferma infatti che la scrittura è un *pharmakon* che, in quanto tale, può avere effetti tanto benefici quanto dannosi: essa aiuta a conservare e a trasmettere l'informazione, anche se non aumenta la memoria e la capacità di ragionare di chi la usa, e non è uno strumento interattivo (275c e segg.).

Ma nel dialogo, Socrate fa anche due discorsi orali, il primo dei quali, pronunciato a capo coperto, viene poi sottoposto da Socrate stesso a una severa critica scientifica. E attraverso il mito delle cicale formula un feroce attacco ai poeti, il cui suono è ripetitivo e non dice nulla (258e-259d).

Platone distingue dunque tra *hypòmnesis*, il bagaglio di nozioni che si possiedono, e *anamnesis*, la loro interconnessione sistematica, secondo un senso unitario e coerente. La componente nozionistica (*hypòmnesis*) può diventare conoscenza scientifica solo se elaborata criticamente attraverso l'*anamnesi*s e in modo interattivo. Mettendo al centro il *logos* filosofico, il filosofo critica la politica culturale dei sofisti, da una parte, e della cultura tradizionale (orale) dall'altra. Il *logos* filosofico può dunque esplicarsi attraverso diversi mezzi, l'importante è che tali mezzi siano usati con scienza, poiché il modo in cui si fa scienza e la si comunica è una questione politica. Un tema, questo che sarà ripreso più volte nella storia del pensiero occidentale. Basti pensare ancora una volta al dibattito illuminista tedesco dell'ultimo quarto del Settecento su stampa, diritto d'autore e scienza, cui prese parte lo stesso Kant[3].

Questo libro si propone dunque di affrontare il rapporto tra la scienza e la rete cercando risposte a una triplice domanda: la rete consente il libero esercizio dell'uso pubblico della ragione? Entro quali limiti? E quali possibilità possono essere esplorate, in tal senso?

[2] Platone, *Fedro*, (tr. it. di P. Pucci) Laterza, Roma-Bari 1966-2000.

[3] Si veda su questo la preziosa raccolta R. Wittman (a cura di), *Nachdruck und geistiges Eigentum*, Kraus International Publications, München 1981.

Il primo capitolo funge da premessa, ed è dedicato a ricostruire le circostanze in cui ha avuto origine il sistema moderno di comunicazione della scienza, dalla nascita del *peer review* e delle riviste scientifiche fino al sorgere degli indici bibliometrici di valutazione, per mettere in luce gli effetti di tali strumenti sul modo in cui si costruisce lo statuto di sapere scientifico. In *The Bias of Communication*, Harold A. Innis ha mostrato come «un monopolio o un oligopolio della conoscenza si costruiscono sino al punto in cui l'equilibrio viene disturbato»[4]. Osserveremo dunque come oggi la rivoluzione mediatica in atto stia cambiando le carte in tavola e rompendo parte di quegli equilibri.

Il secondo capitolo ricostruisce quindi la storia del World Wide Web. La ragnatela globale dell'informazione è stata pensata sin dal principio in analogia con il sistema di comunicazione scientifica tradizionale, ma si basa su principi architettonici diversi, primo tra tutti l'universalità del sistema. Vale a dire che chiunque può pubblicare sul Web senza la necessità di alcun filtro preliminare, una caratteristica che dipende da scelte tecniche, oltre che filosofiche.

> Le strutture sociali del Web dipendono dall'ingegneria che sta sotto alla sua architettura di alto livello. Comprendere la relazione tra umanità e tecnologia e le implicazioni, per la società umana, dell'essere animali che fanno uso di strumenti, è stato un tratto distintivo di gran parte della riflessione filosofica, politica e sociale dall'Illuminismo in avanti, per esempio nell'opera di Marx e di Heidegger. Il Web riflette la nostra vita intellettuale e sociale, ma è anche stato specificamente costruito per essere uno strumento[5].

Un'analisi degli aspetti tecnici più profondi, accompagnata da uno sforzo di tradurli in termini comprensibili agli studiosi di scienze umane e sociali, risulta dunque essenziale per coglierne le implicazioni filosofiche e socio-politiche. Ma il Web ha introdotto altre innovazioni radicali che trasformano il modo di fare scienza in direzioni affatto nuove e rivoluzionarie, e che si scontrano con alcuni postulati della comunicazione scientifica moderna, violando molti dei principi base delle scienze bibliografiche e rendendo difficile certificare la qualità e l'autenticità di un documento. Perciò, nel terzo capitolo si indicano nelle recenti evoluzioni della ragnatela globale, sia da un punto di vista tecnico e filosofico (Web semantico), sia su un piano sociale e culturale (Web 2.0), possibili soluzioni a quei problemi.

[4] H.A. Innis, *Le tendenze della comunicazione*, SugarCo, Milano 1982 (ed. orig. 1951), p. 26.

[5] T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, *A Framework for Web Science*, Foundations and Trends in Web Science, 1, 1, 2006, <http://eprints.ecs.soton.ac.uk/13347/>, p. 82.

Infine, l'ultimo capitolo considera l'impatto delle nuove tecnologie sulla comunicazione scientifica, sia nelle esperienze più conservatrici, che si propongono di adattare gli strumenti tradizionali ai nuovi mezzi, sia in quelle che si poggiano sulla filosofia alla base di Internet, del Web e del movimento per la libertà del software e che mirano a fare della conoscenza un bene comune. «Tra le sfide che la rete pone alla formazione universitaria non può essere ignorato il compito di mantenere alle conoscenze che essa permette di diffondere a fini educativi lo statuto di *bene pubblico globale*»[6]. Una sfida che le università, i centri di ricerca e i ricercatori tutti sono chiamati a raccogliere, come la storia di Alice, nelle conclusioni, cerca di mostrare.

* * *

Desidero in primo luogo ringraziare gli studenti che hanno frequentato i corsi e i seminari che ho tenuto, dal 2004 ad oggi, presso la Facoltà di Scienze politiche dell'Ateneo pisano. Il loro entusiasmo e le loro osservazioni sono stati uno stimolo prezioso. Ringrazio inoltre Alessandro Breccia, Francesca Carpita, Francesca Menchelli-Buttini, Chiara Pasquinelli, Ilaria Possenti, Irene Psaroudakis, ricercatori non strutturati dell'Università di Pisa, che nell'inverno 2009 hanno partecipato a un corso ospitato dal CISIAU della Facoltà di Lettere pisana.

Un pensiero particolare va alla professoressa Elisabetta Marini, per avermi invitato a partecipare al Corso del Master in comunicazione della scienza presso l'Università di Cagliari, e al professor Mario Aldo Toscano, che ha ospitato un analogo ciclo di lezioni rivolto agli studenti del corso di Dottorato in Storia e sociologia della modernità del Dipartimento di Scienze e politiche e sociali di Pisa.

La mia gratitudine va anche ai membri dell'Azione COST A32 "Open Scholarly Communities on the Web", in particolare Paolo D'Iorio, Hans Walter Gabler e Matteo D'Alfonso, per i fruttosi scambi di questi anni.

Questo libro non sarebbe stato concepito senza la guida severa e l'allegria dissacrante di Michele Barbera, che mi ha iniziato allo studio dell'informatica e nei cui confronti ho contratto il debito maggiore. Sono inoltre molto grata a Maria Chiara Pievatolo, che ormai da dieci anni costituisce un insostituibile punto di riferimento.

Infine, grazie a Emanuela, Francesca, Irene, Massi, il calore della cui amicizia mi ha scaldato anche nei momenti peggiori. E a C., che c'è sempre.

[6] L. Gallino, *Tecnologia e democrazia. Conoscenze tecniche e scientifiche come beni pubblici*, Einaudi, Torino 2007, p. 229.

CAPITOLO 1

IL SISTEMA TRADIZIONALE DI COMUNICAZIONE NELLA SCIENZA

1. La nascita della comunicazione scientifica moderna

Gli elementi che hanno caratterizzato il sistema tradizionale di comunicazione in campo scientifico, un sistema che è rimasto per lo più inalterato fino all'ultimo quarto del secolo scorso, si definirono in Europa, e in particolare in Inghilterra, tra la fine del diciassettesimo secolo e i primi decenni del successivo. Se l'universalità della scienza è il risultato di un lungo processo di transizione, fu in Inghilterra, grazie soprattutto alla diffusione che la stampa ebbe in quel preciso contesto storico, culturale e politico[1], che tale processo raggiunse una forma definita e pressoché definitiva: presso la Royal Society di Londra nacque la prima rivista scientifica e si delinearono le pratiche di accreditamento scientifico tutt'oggi in vigore, in particolare il processo di *peer reviewing*[2]; nella dialettica del rapporto tra la gilda degli stampatori e dei librai e gli autori si definirono inoltre i concetti di *copyright* e di autore come soggetto di diritti.

Perciò, tentare di ricostruire il rapporto che esiste tra la diffusione della stampa e l'evoluzione del discorso scientifico - sia per quanto riguarda il merito, vale a dire la costruzione della verità del sapere; sia nel metodo, attraverso un meta-discorso sui mezzi di comunicazione -, è es-

[1] Lo storico Adrian Johns ha in particolare ricostruito il processo tramite cui i testi stampati, in primo luogo scientifici, sono divenuti portatori di verità (cfr. A. Johns, *The Nature of the Book. Print and Knowledge in the Making*, Chicago University Press, Chicago 1998). Al centro della sua analisi è la costruzione del credito e della fiducia (*trust*), un elemento chiave nel creare conoscenza. I processi tramite i quali si decide a chi credere e a che cosa dare credito sono due problemi gemelli, che spesso sono stati combinati in uno solo. Gli storici della stampa si sono interrogati sul fondamento della convinzione che la cultura abbia reso possibile stabilire un sapere vero nella società moderna. Johns, contro altri (in particolare Einsenstein e McLuhan), sostiene che la creazione della “verità” è estrinseca alla stampa in sé, e ricostruisce come vi è stata inglobata.

[2] Secondo David A. Kronick, la pratica del *peer reviewing* sarebbe in realtà stata definita presso la Royal Society of Edimburgh nel 1731 (*Peer review in 18th century scientific journalism*, «JAMA», 263, 10, 1990). Mi paiono tuttavia più attendibili al riguardo le ricostruzioni di Adrian Johns e di Mario Biagioli (*From Book Censorship to Academic Peer Review*, «Emergences», 12, 1, 2002, pp. 91-118).

Francesca Di Donato, *La scienza e la rete : l'uso pubblico della ragione nell'età del Web*,
ISBN 978-88-8453-494-1 (print) ISBN 978-88-8453-500-9 (online)

senziale a comprendere come si è configurato il sistema della comunicazione scientifica tradizionale, un sistema che proprio di recente è stato sconvolto dall'avvento di Internet e del Web, e le cui convenzioni formali fino a quel momento valide e riconosciute risalgono alla fine del '600.

Dalla seconda metà del diciassettesimo secolo Londra, sede della *Stationers' Company*, la corporazione degli stampatori e dei librai riconosciuta legalmente dalla regina Mary nel 1557 «to oversee the 'art and mystery' of printing», divenne il centro del mercato librario. Formalmente, tutti coloro che erano coinvolti nella pubblicazione di libri, riviste, giornali, rientravano nella categoria di *Stationer*, cioè di membri della compagnia che comprendeva i ruoli distinti e articolati di *bookseller*, *printer*, *wholesaler*, *publisher*, *editor* e *compositor*. L'attività degli *Stationers* condizionava fortemente la pratica scientifica: il significato e la comunicazione del sapere di ogni tipo dipendevano sempre più dalla stampa, ed era attraverso l'azione degli *Stationers* che i materiali stampati vedevano la luce e raggiungevano il loro pubblico. In questo senso, si può affermare che il sapere dipendeva dagli *Stationers*[3]. Lo stesso Locke denunciava il fatto che

> il monopolio degli autori classici tenuto dalla Company of Stationers significava che essi erano "scandalosamente male stampati, sia per quanto riguarda i caratteri, la stampa e la correttezza, e a mala pena viene prodotta un'edizione tollerabile", "... la nostra stampa è così sgradevole e tuttavia così cara in Inghilterra"[4].

Nei distretti della capitale del Regno Unito avveniva lo scambio dei libri (anche continentali); fu così che i distretti dedicati al commercio librario giunsero a delineare una specie di università *de facto*: come scriveva Thomas Sprat, autore della storia della Royal Society, «not only the best *Natural*, but the best *Moral* Philosophy too, may be learn'd from the shop of *Mechanicks*»[5].

In questa prima fase, nessuna stamperia londinese poteva permettersi di specializzarsi nella sola letteratura scientifica, che aveva un mercato molto ristretto ed era un'intrapresa ad alto rischio. Un esempio illustre fu l'*English Atlas* di Pitt, la cui pubblicazione, nonostante le ottime credenziali (e il patrocinio scientifico della Royal Society) si arrestò al primo volume degli undici previsti. Per pubblicare un'opera, agli autori (fossero essi Newton, Boyle, Hobbes o Flamsteed) erano necessari ingenti finan-

[3] A. Johns, *The Nature of the Book*, cit., p. 60.

[4] L. King (a cura di), *The Life and Letters of John Locke*, Garland, New York, 1984 (ed. orig. 1864), pp. 204-07; cit. da H.A. Innis, *Le tendenze della comunicazione*, cit., p. 165.

[5] Citato da A. Johns, *The Nature of the Book,* cit., p. 74. Col termine *Mechanick* si intendevano gli stampatori. In Europa, che il sapere filosofico vivesse nei negozi così come nei libri e nelle accademie era una idea nota e condivisa.

ziamenti, e i libri scientifici erano una merce cara, sia da produrre che da comprare. Anche per questo, la prima e più importante difficoltà per un autore consisteva nel persuadere uno *Stationer* a stampare[6]. Fu certamente anche a causa di tali difficoltà che i filosofi naturali della Royal Society, autori di libri destinati a un pubblico erudito e a un mercato di nicchia, inaugurarono e sperimentarono soluzioni che consentirono loro di crearsi uno spazio di autonomia nel mondo degli *Stationers*[7].

La comunicazione era essenziale per la Royal Society, che si autoproclamava "parlamento" intendendosi con ciò rappresentativa della nazione, ma si apriva al pubblico (internazionale) dei lettori tramite le pubblicazioni, che godevano di considerevole fama anche oltre la Manica. Il bisogno di diffondere i risultati era importante per la filosofia sperimentale al pari dell'esperimento, e la prima condizione per essere un filosofo era la prontezza a comunicare[8].

L'accademia inglese avviò dunque un processo che le consentì di affrancarsi dalla Compagnia degli *Stationers* quando ottenne il privilegio di stampare opere scientifiche e l'autonomia di scegliere i suoi librai e stampatori[9]. Infine, il consiglio della Royal Society stabilì che nessun libro potesse ottenere una licenza se non fosse stato prima approvato da almeno due membri del consiglio stesso, per assicurare che esso non fosse in alcun modo contrario ai progetti e alle opere della società (una prima forma di *peer review*)[10].

Fu in tale contesto che nacque un genere di pubblicazioni che si offriva come veicolo regolare per pubblicare articoli che avessero superato la valutazione della Società: il periodico scientifico. Come per il *Journal des sçavans*, fondato in Francia da Denis de Sallo, il primo numero di *The Philosophical Transactions* apparve nel 1665 a cura di Henry Oldenburg[11], segretario, responsabile della corrispondenza e custode dei regi-

[6] Una soluzione che trovò seguito nella Royal Society, e anche altrove in Europa, fu l'autofinanziamento da parte degli autori (e tramite gli abbonamenti dei lettori). L'ostacolo principale era trovare sottoscrittori che si fidassero dell'iniziativa editoriale. Cfr. su questo anche la proposta del poeta Friedrich Gottlieb Klopstock, che ebbe notevole seguito anche presso altri illuministi tedeschi (M. Woodmansee, *The Genius and the Copyright: Economic and Legal Conditions of the Emergence of the 'Author'*, «Eighteenth-Century Studies», 17, 4, 1984, 425-448, in particolare pp. 440-41, <http://www.compilerpress.atfreeweb.com/Anno%20Woodmansee%20Genius%20&%20Copyright.htm>).

[7] Cfr. T. Sprat, *History of the Royal Society*, Martyn, London 1667; T. Birch, *The History of the Royal Society of London,*1756, reprinted Johnson, New York 1968.

[8] A. Johns, *The Nature of the Book*, cit., p. 472.

[9] Nel 1660-61 la Royal Society scelse gli stampatori John Martyn e James Allestry, con i quali fu stabilita una rigida convenzione.

[10] Cit. in A. Johns, *The Nature of the Book*, cit., p. 494.

[11] Oldenburg gestiva la nuova rivista con grande indipendenza dal punto finanziario ed editoriale e *The Philosophical Transactions* fu a lungo identificata con la sua persona.

stri della Society[12]. Il tentativo di Oldenburg era probabilmente mirato a estendere il valore del registro dell'accademia inglese al di fuori della Society stessa, e il ruolo di editor delle *Philosophical Transactions* e di guardiano (*gatekeeper*) del registro si intersecarono costantemente.

La nascita di questo nuovo genere di pubblicazioni introdusse cambiamenti di notevole rilievo sia nel mercato editoriale, sia nelle forme della comunicazione scientifica[13]. Le riviste posero problemi pratici e legali che non esistevano coi libri: richiedevano licenze aperte e illimitate, avevano bisogno di numerosi valutatori e necessitavano di una grande quantità di contenuti. Anche per questo, si aprirono al contributo degli studiosi esteri favorendo l'internazionalizzazione della scienza. Lentezza delle pubblicazioni, plagio, conflitti personali e politici oltre che scientifici, comportamenti fraudolenti, bassa *expertise*, eccessivo conservatorismo cominciarono allora ad essere denunciati come fenomeni negativi.

Anche la nascita del *peer review* in senso proprio, un momento decisivo per la costruzione del credito del sapere scientifico e della conseguente fiducia da parte del pubblico, viene fatta coincidere con quella del primo periodico scientifico, le appena citate *Philosophical Transactions*. La Royal Society dovette infatti confrontarsi direttamente con l'elaborazione di convenzioni sicure per la produzione, la manipolazione e la ricezione di oggetti scritti e stampati che contenevano teorie di filosofia naturale al fine di difendere il proprio privilegio, che la Corona avrebbe potuto revocare in qualunque momento. Fu in tale contesto che i membri dell'accademia inglese dettero vita alle prime pratiche di revisione da parte di pari: presentazione e "lettura attenta" (*perusal*), più una forma di registrazione molto simile a quella che avveniva nella gilda degli *Stationer*. La presentazione era un atto pubblico, spesso mediato dal segretario della società, che divenne obbligatorio dal maggio 1661: ogni membro che pubblicava un'opera doveva donarne una copia alla Royal Society; in

[12] Se il periodico francese e la rivista inglese sono stati spesso equiparati, «la pubblicazione francese rifletteva in realtà il modello degli scambi epistolari manoscritti, in qualche modo pettegoli e orientati alla notizia, che erano così tipici della Repubblica delle Lettere; come tale, è più vicina a qualcosa come lo *Scientific American*, piuttosto che a una rivista accademica moderna, e perciò appare saldamente radicata nell'arte emergente del giornalismo scientifico. Sebbene la rivista pubblicasse occasionalmente articoli originali, essi apparivano come una espressione particolare di notizie fra altri tipi di notizie. Di contro *Philosophical Transactions*, sebbene trattasse anch'esso di nuova informazione, mirava in realtà a creare un registro pubblico di contributi originali alla conoscenza. In altri termini, la pubblicazione parigina andava dietro alla novità, mentre la rivista londinese contribuiva a corroborare l'originalità. In questo consiste la differenza significativa (e profonda) fra i due periodici». J.-C. Guédon, *Per la pubblicità del sapere. I bibliotecari, i ricercatori, gli editori e il controllo dell'editoria scientifica*, PLUS, Pisa 2004 (ed. orig. 2001), p. 18.

[13] Per una rassegna sulla storia dei periodici in Europa, cfr. B. Dooley, S. Baron (a cura di), *The Politics of Information in Early Modern Europe*, Routledge, New York 2001.

seguito, il dono meritava una risposta: di solito un *fellowship*. La "lettura attenta" consisteva in un'analisi approfondita dell'opera da parte di specialisti della medesima accademia. Il giudizio dei valutatori doveva essere comunicato all'autore e le conclusioni di questi ultimi non dovevano espressamente essere rese pubbliche.

Dopo la "lettura attenta" le opere venivano inserite in un registro che ricalcava il registro delle copie della compagnia degli *Stationer*, un volume scritto a mano e custodito dal *clerk* della compagnia, che offriva a colui che registrava un testo un diritto perpetuo sulla base delle convenzioni della gilda stessa[14]. La registrazione, affermava Boyle, avveniva «[to] secure [authors] against the usurpations, not the industry or out-doings of others»[15], dunque per stabilire la priorità di un individuo in una scoperta scientifica. Dei due registri esistenti, un libro era dedicato alle lettere, l'altro alle teorie (ipotesi), e l'accesso (a entrambi) era limitato. Si osservi inoltre che l'accademia inglese si riservava il diritto di modificare le opere prima di registrarle. La procedura suggerita da Oldenburg per assicurare la paternità delle invenzioni ai loro autori e per prevenirne l'usurpazione divenne così un protocollo. Tuttavia, il ruolo della Royal Society fu sempre fragile e l'autorità dei registri non si estese al di fuori della comunità accademica.

Nella pratica di *peer review* messa in atto dal periodico inglese, era la rivista intera a essere sottoposta a giudizio, e non come oggi i singoli articoli. Almeno al principio, inoltre, gli stessi contenuti erano difficilmente riconducibili a un autore altro da Oldenburg, che, pur ricevendo manoscritti anche dall'estero, pubblicava più parafrasi di suo pugno e brevi *report* anonimi che testi originali. Una particolare caratteristica del sistema di comunicazione scientifica inglese era il fatto che il processo di *peer review* si applicava ai prodotti domestici (e avveniva all'interno della Royal Society), e la censura interessava invece i libri importati dall'estero. I "pari" dunque erano tra loro connazionali[16]. Grazie alle pubblicazioni, le accademie reclutavano nuovi eruditi, ampliavano le loro reti e incoraggiavano l'istituzione di nuove accademie. Nel rapporto tra scienziati e politica, il *peer review* era importante perché la credibilità (e l'indipendenza) dell'istituzione dipendevano dalla credibilità dei testi pubblicati e dalla rete di scambi filosofici dei membri dell'accademia. Perciò, il *peer review* fu in principio uno strumento prudenziale, e assai difficilmente

[14] Una volta registrata da uno *Stationer* (procedura che avveniva dietro il versamento di 6 pence), una copia poteva essere venduta, scambiata, suddivisa in parti ma la registrazione era il documento che aveva autorità sul diritto consuetudinario. Il sistema comportava inoltre che per verificare lo status di un titolo fossero necessarie considerevoli ricerche. A. Johns, *The Nature of the Book*, cit., pp. 213-16.

[15] *Ivi*, p. 484.

[16] Fu invece a Parigi che ebbe inizio la valutazione a distanza, che anticipa la pratica del *blind peer review*. Cfr. M. Biagioli, *From Book Censorship to Academic Peer Review*, cit.

opere innovative passavano il vaglio della Royal Society e venivano date alle stampe.

In generale, la costruzione del credito delle pubblicazioni non avvenne in modo omogeneo. Nel mentre, tuttavia, le pubblicazioni scientifiche divennero un requisito necessario per appartenere all'accademia – e le pubblicazioni, comparendo sulle riviste accademiche dell'istituzione, davano prestigio all'istituzione stessa. In questo modo, i manoscritti divennero portatori di credito. E per il funzionamento del meccanismo, era fondamentale che queste "banconote accademiche" fossero stampate[17].

Tra Sei e Settecento il *peer review* estese la sua giurisdizione. Nato come tecnica disciplinare e interna ad accademie specifiche, il suo uso fu poi allargato alla valutazione dei risultati degli eruditi in generale. Se non è facile distinguere i diversi momenti (presentazione, "lettura attenta", registrazione, pubblicazione) della valutazione, possiamo osservare che fu a causa dello stretto legame tra le società scientifiche e i loro periodici che il *peer review* divenne parte integrante nella pubblicazione delle prime riviste e, per estensione, un protocollo di pubblicazione nella Repubblica delle Lettere del 1700. In seguito, si generalizzò e fu istituzionalizzato come pratica tacita ma universalmente riconosciuta nell'accademia.

Dal 1662 fu il Press Act a stabilire che alla Royal Society spettasse, per prerogativa regia, il privilegio di licenziare i testi per la stampa. Il regime del privilegio si intrecciava a quello delle licenze: i licenziatari erano autorità a cui la Corona attribuiva il compito di vagliare il contenuto dei testi che potevano essere stampati, compito che fino al 1643 era assegnato in prevalenza a ecclesiastici, e che a partire da quella data venne attribuito a una nuova categoria di esperti tra cui, nel 1662, rientrò la Royal Society. Si osservi che il ruolo di licenziatario era più importante di quello di autore, soprattutto in termini di responsabilità nei confronti dei contenuti di un testo; a prova di ciò sia considerato il fatto che, dal 1677, il nome del concessionario della licenza doveva comparire nelle voci dei testi incluse nel registro della compagnia degli *Stationers* (disposizione che fu "resa obbligatoria per ordine regio" nel 1685), mentre il nome dell'autore spesso veniva omesso. Elemento, questo, che può essere compreso considerando che l'attribuzione di privilegi e di licenze equivaleva a una forma di censura assai efficace e potente, che garantiva un controllo assoluto da parte del potere politico sui materiali stampati.

Un modo affinché un autore potesse ottenere una licenza era persuadere un importante *gentleman* (ad esempio un *Master Printer*) a scrivere alla corte chiedendo il permesso di stampare un'opera sotto il suo brevetto. A metà del Seicento lo stato e gli *Stationers* avevano sviluppato un insieme complesso di procedure per stabilire e regolare la proprietà dei libri, al cui centro stava il registro delle copie, su cui venivano inseriti titoli, date e, appunto, il nome di chi aveva concesso la licenza. Ma il registro

[17] M. Biagioli, *From Book Censorship to Academic Peer Review*, cit., p. 20.

non era esplicitamente protetto dalla legge; pertanto, “proprietà letteraria” è un termine anacronistico e sbagliato per il tempo.

Così, il registro delle copie (di cui a partire dal 1662 si dotò anche la Royal Society), le licenze e i privilegi reali divennero elementi fondamentali per la costruzione del credito dei testi stampati.

> Il regime del privilegio resistette anche all'indebolimento del potere monarchico determinato dalla prima e dalla seconda rivoluzione inglese: quando nel 1641 fu abolita la Star Chamber, che assicurava sia le licenze di stampa sia il monopolio della Stationers' Company, i membri della corporazione presentarono il pamphlet *The Humble Remonstrance of the Company of Stationers to the High Court of Parliament* (1643), che stigmatizzava le Province Unite dei Paesi Bassi per la mancanza di regolamentazione che impoveriva gli stampatori e impediva loro di operare al servizio dello stato. Il Licensing Order, emanato dal Parlamento nel 1643, reintrodusse un regime di monopolio e di censura analogo a quello precedente. Dopo la restaurazione del potere monarchico, il Licensing Act (1662) ribadì questa disciplina, che durò fino al 1695 - anno in cui il Licensing Act non fu più rinnovato[18].

Fu a partire dal 1695, e poi dal 1710, che la sede decisionale in materia di privilegi di stampa e poi di proprietà intellettuale passò dalla corte della gilda a quella del *common law*, un passaggio decisivo per la nascita dell'attuale sistema di *copyright*, che ha trasferito i diritti sulla “copia” dal concessionario della licenza all'autore. Fu così che

> nel 1710 il parlamento britannico approvò la prima legge europea sul copyright, lo *Statute of Anne*. Per la prima volta, l'autore, in luogo dello stampatore, è riconosciuto come titolare originario del monopolio sulla riproduzione del proprio lavoro. Questo monopolio, però, non è più perpetuo, bensì temporaneo: il termine è di 21 anni per le opere già pubblicate al momento dell'entrata in vigore dello statuto e, per tutte le opere uscite dopo, di 14 - raddoppiabili solo con un atto esplicito di volontà dell'autore. Il termine di 14 anni era stato ripreso da quello dei brevetti sulle invenzioni stabilito dallo *Statute of Monopolies* del 1623: in entrambi i casi, il monopolio era inteso non come una proprietà di diritto naturale, ma come una costruzione politica[19].

[18] M.C. Pievatolo, *La comunicazione del sapere. La questione del diritto d'autore*, «Bollettino telematico di filosofia politica», 2007-2008, <http://bfp.sp.unipi.it/dida/fpa/ar01s02.html>.

[19] *Ibidem*.

2. I sistemi di comunicazione della scienza e la valutazione del sapere

In letteratura i modelli di scambio del sapere sono molti e includono, in generale, diverse funzioni. John Willinsky[20] ne individua tre scopi principali, vale a dire:

1. lo *sviluppo* di conoscenze innovative;
2. la *valutazione* della qualità e la discriminazione tra sapere valido ed eccellente e conoscenze false e ininfluenti, e infine
3. la *disseminazione* e l'effettiva efficacia delle conoscenze proposte al pubblico.

Van de Sompel e altri ripropongono un modello presentato da Rosendaal e Geurts[21], basato su cinque momenti fondamentali:

> la *registrazione*, che permette l'esercizio di diritti di precedenza su una scoperta scientifica; la *certificazione*, che stabilisce la validità di una scoperta una volta registrata; la *presa di coscienza*, che consente agli attori del sistema scientifico di prendere coscienza di nuovi titoli e scoperte; l'*archiviazione*, che ne garantisce la registrazione nel tempo; e infine la *ricompensa*, che premia gli attori per la loro prestazione nel sistema di comunicazione sulla base di metriche che derivano dal sistema stesso[22].

Al centro di ciascuno di essi si trovano le pubblicazioni, nella forma di articoli su periodici o di monografie. Metro di valutazione universale del valore scientifico, riconosciuto sul piano nazionale e in campo internazionale, la pubblicazione dei risultati scientifici è diventata nel corso dei secoli un elemento cardine della comunicazione scientifica, in cui le riviste hanno assunto un ruolo predominante. John Ziman definisce «comunicazione scientifica primaria» ogni contributo originale dato alla conoscenza, pubblicato normalmente nella forma di saggio o di articolo su una rivista dedicata a una specifica disciplina[23]. In parallelo alla specializzazione disciplinare, un numero sempre crescente di periodici ha visto la luce[24]. Inoltre, le pubblicazioni sono da sempre legate al problema della

[20] J. Willinsky, *Proposing a Knowledge Exchange Model for Scholarly Publishing*, «Current Issues in Education», 3, 6, 2000, pp. 1-6.

[21] H. Roosendaal, P. Geurts, *Forces and Functions in Scientific Communication: an Analysis of their Interplay,* «Cooperative Research Information Systems in Physics», 1997, <http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html>.

[22] H. Van de Sompel, S. Payette, J. Erickson, C. Lagoze, S. Warner, *Rethinking Scholarly Communication: Building the System that Scholars Deserve*, «D-Lib Magazine», 10, 9, 2004, <http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html>.

[23] J. Ziman, *Il lavoro dello scienziato: gli aspetti filosofici e sociali della scienza e della tecnologia*, Laterza, Roma-Bari 1987 (ed. orig. 1984), p. 80.

[24] Circa un centinaio fino al 1800, un migliaio cinquant'anni dopo, 10.000 intorno al 1900, nel 1965 D.J. De Solla Price ne stima circa 35.000. Cfr. S. Sacchi, *Comuni-*

valutazione, un momento fondamentale che interessa il pubblico (in termini di ricaduta in investimenti pubblici, di linee di sviluppo, di sostenibilità della ricerca), finanziatori e investitori pubblici e privati, università e centri di ricerca e singoli ricercatori (in quanto determinante per l'accesso all'accademia e per fare carriera al suo interno).

Il problema della valutazione nell'università chiama in causa diversi aspetti: si valutano la ricerca e la didattica, e si valutano tanto i singoli quanto le istituzioni. La valutazione dei risultati avviene inoltre su un piano tanto nazionale quanto internazionale (si pensi, ad esempio, all'attribuzione di fondi di ricerca su base europea, uno strumento sempre più utilizzato per finanziare la ricerca di eccellenza dei singoli stati).

Infine, la valutazione della scienza avviene secondo criteri di natura diversa: qualitativi, primo tra tutti il *peer review*, e, più di recente, quantitativi. Si tratta di criteri che vengono applicati in modo differente nelle diverse discipline e che solitamente vengono ricondotti a due macro-aree disciplinari, scienze, tecnologia e medicina (STM) da una parte, e dall'altra scienze umane e sociali (SSH). Nelle prime, gli articoli su rivista sono il principale strumento di accreditamento di uno studioso, nelle seconde sono invece le monografie a giocare un ruolo predominante. Nonostante l'efficacia di tale suddivisione, esistono al suo interno vere e proprie isole disciplinari in cui vigono meccanismi diversi. A titolo di esempio si consideri che la filologia classica e l'economia, che appartengono entrambe alle SSH, si basano su sistemi di pubblicazione e di valutazione più distanti di quanto non valga per la stessa economia e le scienze biomediche. Si deve inoltre osservare in generale che, mentre il sistema delle pubblicazioni nelle STM è stato analizzato e studiato[25], il mercato delle pubblicazioni nelle SSH resta per lo più un campo oscuro, non esistendo nessuno studio comparabile a quelli disponibili per le discipline più ricche e maggiormente finanziate[26]. Tuttavia, al di là dei tratti specifici delle diverse discipline, la scienza è fatta di attività comuni (che vanno dal ruolo e dalla posizione in istituzioni nazionali e internazionali, al controllo delle pubblicazioni, alla capacità di ottenere finanziamenti, posti e

cazione scientifica e open access. Problematiche sociali e tecnologiche nell'applicazione in Italia, tesi di laurea, 2003-04, <http://dspaceunipr.cilea.it/bitstream/1889/359/2/comunicazione_scientifica_e_open_access_simone_sacchi.pdf>.

25 Cfr. OECD, *Report on scientific publishing of 2004-2005,* DSTI/ICCP/IE(2004)11/FINAL,<http://www.oecd.org/dataoecd/42/12/35393145.pdf>; M. Dewatripont et al., *Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe*, Directorate - General for Research, European Commission, 2006, <http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf>.

26 Per le SSH manca ad esempio uno studio comparabile a quelli citati nella nota precedente. Fa eccezione un recentissimo articolo di Maria Cassella, che offre dati e spunti interessanti. Cfr. M. Cassella, *La valutazione della ricerca nelle scienze umane*, «quaderni CNBA», 11, 2009.

promozioni) in cui gli scienziati competono per un monopolio sull'autorità scientifica. Di contro, l'autorità scientifica è costruita tanto come capacità tecnica (vale a dire scientifica) quanto come potere sociale, in cui l'accesso alle pubblicazioni e il controllo su di esse sono parte essenziale[27].

La natura competitiva della scienza è data per assunta da molti, ma la sua evoluzione in ciò che oggi è sempre più spesso definita come una struttura di potere oligarchica[28] è poco presente nel dibattito, e non viene messa in questione. Per discuterne i fondamenti, un'analisi del sistema di comunicazione e delle modalità di valutazione risulta pertanto essenziale.

Si è accennato al fatto che la valutazione avviene secondo criteri qualitativi e quantitativi, la cui scelta influenza efficacia e credibilità dei risultati del processo.

Nel sistema tradizionale di comunicazione scientifica, fanno parte dei primi il *peer review*, che è antecedente alla pubblicazione, e le recensioni, ad essa successive. Si è osservato come, sin dalle origini, il valore scientifico si costruisse tramite il giudizio di pari e la Royal Society regolasse il proprio privilegio con la norma secondo cui nessun libro poteva ottenere una licenza se non fosse stato prima approvato da almeno due membri del suo consiglio, per assicurare che esso non fosse contrario ai progetti e alle opere della società. Lungi dal rappresentare un concetto giuridico, il *peer review* è divenuto così il filtro per ammettere nuovi membri all'interno della società delle lettere ma anche per decidere quali opere pubblicare. Il *peer review*, nato nel mondo del privilegio e delle accademie, nasce dunque come una garanzia di qualità che è strettamente connessa alla tecnologia della stampa. Garantendo una scienza di qualità, esso serve a preservare il contratto tra la comunità scientifica e il potere politico, non senza compromessi. Nelle SSH il *peer review* resta, ad oggi, il principale modo di valutazione. Ciononostante si tratta di un metodo soggettivo che, in quanto tale, è considerato da più parti discrezionale.

Dall'inizio del diciannovesimo secolo «è andata costituendosi un'estesissima letteratura *secondaria*, formata da *bibliografie*, *periodici di 'riassunti'*, *compilazioni di dati, recensioni* e altre pubblicazioni regolari, che catalogano o passano in rassegna i contenuti della letteratura primaria», strumenti che, nel corso del Novecento, sono stati computerizzati[29]. Per rispondere alle esigenze di documentaristi e bibliotecari, che si trovavano

[27] Cfr. P. Bourdieu, *La spécificité du champ scientifique et les conditions sociales du progrès de la raison*, «Sociologie et sociétés», 7, 1, 1975, pp. 91-118 e J.-C. Guédon, *Open Access. Contro gli oligopoli nel sapere* (a cura di F. Di Donato), ETS, Pisa 2009.

[28] B. Whitworth, R. Friedman, *Reinventing Academic Publishing online. PART I. Rigor, Relevance and Practice*, «First Monday», 14, 8, 2009, <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2609/2248>.

[29] J. Ziman, *Il lavoro dello scienziato*, cit., p. 81.

a fronteggiare la necessità di disporre dei periodici scientifici di ogni disciplina, a partire dal primo dopoguerra è nata la bibliometria, una scienza che «utilizza tecniche matematiche e statistiche per analizzare i modelli di distribuzione dell'informazione», il cui «successo è dovuto alla ricerca, da parte della comunità scientifica mondiale, di parametri oggettivi per la misurazione, detti indicatori bibliometrici, e la conseguente valutazione, della scienza e degli scienziati»[30]. Si tratta di criteri quantitativi *ex post*, vale a dire indici che servono a calcolare l'impatto di una pubblicazione sulle successive e a stabilire quali riviste siano da considerare irrinunciabili[31]. Nell'adozione di questi sistemi, la crisi del '29 ha segnato uno spartiacque decisivo.

> Negli anni Trenta gli ambienti di ricerca furono fortemente riformati per far fronte al crollo verticale dell'economia americana e le biblioteche, che fino ad allora avevano avuto un budget sufficiente a coprire lo sviluppo delle collezioni, e in particolare gli abbonamenti ai periodici scientifici, si sono ritrovate nell'impossibilità di fornire un servizio adeguato alle sempre crescenti richieste d'informazione[32].

La contrazione dei bilanci, e l'aumento esponenziale del numero di testate, hanno portato alla nascita di strumenti per selezionare le pubblicazioni cosiddette fondamentali. Agli anni '30 risale la legge di Bradford che, come il principio di Pareto noto anche come legge 80/20, ha mostrato come alcune riviste (divise per disciplina) sembrino essere in grado di produrre più articoli interessanti rispetto ad altre, portando all'individuazione di "*core journals*", riviste che sono definite fondamentali in quanto considerate le migliori fonti di informazione in un campo di ricerca determinato[33]. Uno strumento essenziale al fine di individuare la cosiddetta scienza fondamentale sono le citazioni; la scienza, in quanto attività cumulativa e progressiva, è il prodotto dell'interconnessione di opere recenti con opere precedenti.

> Questo fatto è in accordo con la nostra comprensione filosofica della conoscenza scientifica nella forma di una rete strettamente interconnessa di fatti e concetti e con la nostra osservazione sociologica che gli scienziati appartengono a una comunità fortemente interattiva. Il

[30] A. De Robbio, *Analisi citazionale e indicatori bibliometrici nel modello Open Access*, «Bollettino AIB», 47, 3, 2007, pp. 257-288, <http://eprints.rclis.org/11999/>.

[31] Sui limiti di tali indici, si veda OECD, *The measurement of scientific and technical activities: R&D statistics and output measurement in the higher education sector*, Paris, 1990.

[32] S. Sacchi, *Comunicazione scientifica e open access. Problematiche sociali e tecnologiche nell'applicazione in Italia*, cit., p. 20.

[33] S.C. Bradford, *Sources of Information on Specific Subjects*, «Engineering: An Illustrated Weekly Journal», 137, 1934, pp. 85-86.

sistema di comunicazione formale della scienza rende del tutto evidenti queste interconnessioni e interazioni[34].

Tuttavia, il campo creato dalla rete di citazioni non ha una natura uniforme. Osservava infatti Derek J. De Solla Price in uno studio del 1965 che, se le citazioni indicano la natura di un campo scientifico, sul campione studiato da Price si poteva osservare che il 35% degli articoli non veniva mai citato, il 49% aveva ricevuto una sola citazione, il 16% aveva circa 3 citazioni mentre solo l'1% ne aveva ricevute più di sei. «Sono tentato di concludere, scriveva infine, che delle 35.000 riviste correnti una frazione molto ampia dev'essere considerata un mero rumore di sottofondo, molto distante dall'essere centrale o strategico in ognuna delle strisce di cui è fatto l'abito della scienza»[35].

Nel tradurre tali principi in strumenti di valutazione, si è rivelato decisivo l'apporto di Eugene Garfield, che ha trasformato la legge di distribuzione di Bradford nella sua legge di concentrazione creando un indice delle citazioni scientifiche, lo Science Citation Index (SCI), che potesse essere considerato una buona rappresentazione di "*core science*". Lo SCI fu ideato nei primi anni Cinquanta da Garfield in risposta all'esigenza di poter disporre di un sistema bibliografico per la letteratura scientifica in grado di rendere ininfluenti le citazioni non rilevanti (di dati falsi, incompleti e obsoleti). Uno strumento finalizzato a permettere alle biblioteche di scegliere la letteratura di rilievo, ai fini delle loro politiche di acquisto; e pensato perché i ricercatori potessero costruire facilmente bibliografie pertinenti e complete. Non potendo leggere tutto, chi fa ricerca deve poter disporre di criteri di qualità sui quali basarsi: lo scopo di un indice delle citazioni scientifiche sarebbe stato quello di rendere possibile al ricercatore la selezione delle citazioni importanti e degli articoli fondamentali per la sua ricerca. Sostenendo l'importanza di identificare strumenti in grado di rendere più semplice il compito degli storici, per i quali, al crescere della quantità della letteratura scientifica «diviene sempre più difficile separare il grano dal loglio»[36], Garfield pone così al centro della riflessione due elementi essenziali: l'importanza delle citazioni; e il problema della selezione dell'informazione scientifica.

Perché un indice basato sulle citazioni? «Se uno considera i libri le macro-unità del pensiero, e gli articoli di rivista le micro-unità, l'indice delle citazioni si concentra sulle sub-micro-unità (molecolari) del pen-

[34] J. Ziman, *Il lavoro dello scienziato*, cit. p. 83.

[35] D.J. De Solla Price, *Networks of Scientific Papers. The pattern of bibliographic references indicates the nature of the scientific research front*, «Science», 194, 3683, 1965, pp. 510-15, <http://www.garfield.library.upenn.edu/papers/pricenetworks1965.pdf>.

[36] E. Garfield, I.H. Sher, Richard J. Torpie, *The use of citation data in writing the history of science*, 1964, <http://www.garfield.library.upenn.edu/papers/useofcitdatawritinghistofsci.pdf>.

siero», scriveva Garfield. Gli scienziati, spesso, sono interessati a un'idea particolare piuttosto che a una concezione generale, e se appropriatamente pensati e sviluppati, indici di "pensieri" possono essere estremamente utili ai fini della selezione dell'informazione: nel processo della ricerca bibliografica, gli indici per soggetto ricoprono un ruolo minimo, anche se significativo, poiché servono semplicemente da punto di partenza della ricerca. I limiti di questo mezzo, secondo Garfield, sono da riscontrarsi nel fatto che è impossibile creare indici per soggetto che coprano tutte le possibili impostazioni di cui lo scienziato ha bisogno. Gli indici a soggetto sono schematici, e la possibilità di incrementare il numero dei soggetti non elimina la rigidità del mezzo; perciò, neppure una migliore standardizzazione della classificazione sarebbe stata risolutiva. Viceversa, Garfield riscontrò la necessità di nuovi strumenti bibliografici che aiutassero a colmare la distanza tra chi crea il documento, cioè l'autore, e lo scienziato che cerca l'informazione.

Senza nasconderne gli aspetti problematici[37], l'indice delle citazioni viene così descritto come un indice «basato sull'associazione delle idee», in grado di offrire al lettore tutta la libertà d'azione che richiede, rendendo possibile tradurre in pratica un metodo di ricerca modellato sul soggetto. Garfield fece riferimento ad alcuni esempi di indici che funzionavano, già dalla prima metà dell'Ottocento, grazie a sistemi analoghi, e propose di sviluppare un codice di citazioni per la scienza, in cui a ogni articolo fosse associata una doppia serie di numeri, la prima corrispondente alla rivista in cui è stato pubblicato, la seconda scelta arbitrariamente, che identifica l'articolo stesso.

Sotto ciascun codice numerico, "11123s-687" nell'esempio in figura 1 della pagina che segue, vengono ordinati altri codici numerici che rappresentano gli articoli che citano l'articolo in questione, accompagnati da un'indicazione sul tipo di fonte citante. In pratica, il sistema offre una lista completa di tutti gli articoli originali che hanno citato l'articolo in questione, ma limitata a un insieme di riviste chiuso: quelle che appartengono all'indice. È lo stesso Garfield a mettere in luce che l'impatto di un articolo si riferisce a un elenco di pubblicazioni limitato. Perché tale sistema funzioni è infatti necessario stendere un elenco delle pubblicazioni (le riviste) e catalogarle (assegnando a ciascuna un codice, nell'esempio "11123s").

[37] «Un alto numero di citazioni riflette l'impatto, ma può anche non riflettere il valore intrinseco. I dati ottenuti dall'analisi delle citazioni sono sempre relativi, piuttosto che assoluti». E. Garfield, Irving H. Sher, Richard J. Torpie, *The use of citation data in writing the history of science*, cit., p. 10.

Table 1

Index sample based on article by Hans Selye, "General adaptation syndrome" [J. Clin. Endocrinol. '6, 117 (1946)]

The code number for this journal in the World List is 11,123a; the article number is arbitrarily taken as 687 ; and the code number for the article is 11123a-687. The 23 articles that cited Selye's article are listed, followed by A hypothetical citation index entry for Selye's article: R, review article ; A, abstract ; 0, original article.

1. Williams, R. H. : Thyroid & Adrenal Interrelations, 7: 52—57 (1947).
2. Venning, E. H. : Glycogenic Corticoids, 7: 79—101 (1947).
3. Forbes, et al. : 17-Ketosteroids in Trauma and Disease, 7 : 264—288 (1947).
4. Talbot, et al. : Excretion of 11-Oxy-corticosteroids, 7: 331—350 (1947).
5. Castillo, E. B. del, et a!. : Syndrome of Rudimentary Ovaries, 7 : 385—422 (1947).
6. Forsham, P. H., et al. : Pituitary Adrenocorticotropin, 8 : 15—66 (1948).
7. Pincus, G., et al : Rhythm in Biped Excretion, 8: 221—226 (1948).

[...]

Figura 1. Un esempio di sistema di codifica di un articolo[38]

L'esempio nella figura seguente mostra l'elenco delle citazioni di un articolo, e indica le fonti citanti.

Citation Index Entry

11123s-687

____

464-9789(R)
869-3366(R)
1105-9876(A)
1123-4432(R)
a11,123-0752(0)
-0779(0)
-7264(0)
-7331(0)
-7385(0)
-0866(0)
-8221(0)
-9158(0)
-9497(0)
-9529(0)

Figura 2. Una voce dello SCI

[38] Nella figura l'elenco delle fonti citanti si limita a sette voci; per l'elenco completo e una descrizione dettagliata del meccanismo, si veda: E. Garfield, *Citation Indexes for Science. A New Dimension in Documentation through Association of Ideas,* «Science», 122, 3159, 1955, pp. 108-11 («Essays of an Information Scientist», 6, 1983, pp. 468-471),<http://www.garfield.library.upenn.edu/essays/v7p525y1984.pdf>.

Nel medesimo articolo Garfield suggeriva possibili impieghi pratici di un simile indice, che si sarebbe rivelato in particolare molto utile nelle scienze umane, filosofiche e sociali, per tracciare percorsi che ricostruissero l'origine e l'evoluzione di un'idea e nelle ricerche storiche al fine di valutare il significato di un'opera calcolando il suo *impatto* sulla letteratura e sul pensiero successivi. Garfield aggiungeva che un tale fattore d'impatto (*Impact Factor*) avrebbe potuto rivelarsi uno strumento di valutazione più indicativo e significativo rispetto al conteggio assoluto del numero delle pubblicazioni scientifiche. Infine, l'indice delle citazioni avrebbe facilitato la comunicazione tra scienziati e la nascita e lo sviluppo di nuove idee.

Garfield si poneva il problema di gestire grandi quantità di dati, e propose di applicare il proprio sistema a un numero ristretto di riviste, poi denominate *core journals*, cioè le riviste più importanti e prestigiose. Il sistema abbracciava, dalla sua ideazione, un'impostazione centralizzata dell controllo bibliografico – e se i limiti di tale impostazione sono connaturati al mondo della stampa, e un'impostazione distribuita poteva suonare impensabile a causa della limitatezza dei mezzi a disposizione (in primo luogo dall'assenza della rete), questi limiti hanno avuto riscontri negativi di portata ampia.

Lo SCI, «da sistema per risolvere al meglio un problema organizzativo all'interno delle biblioteche (la possibilità di individuare con certezza quali periodici acquistare, in base al loro reale utilizzo) divenne un sistema di valutazione della scienza nel suo complesso»[39], soprattutto grazie all'adozione dell'*Impact Factor* (IF), una forma di misura standardizzata introdotta dall'*Institute of Scientific Information* (ISI) che si calcola dividendo il numero di citazioni che una rivista riceve, in un dato anno, relative agli articoli pubblicati nei due anni precedenti, per il numero di articoli pubblicati in quegli stessi anni. Ad esempio, il fattore d'impatto di una rivista per il 2008 è il numero di citazioni ricevute nel 2008 relative agli articoli pubblicati nel 2007 e nel 2006, diviso per il numero di articoli pubblicati nel 2007 e nel 2006[40].

È stato osservato che l'uso dell'*Impact Factor* «per la valutazione della ricerca scientifica individuale o collettiva, ad esempio, di dipartimenti o gruppi di ricercatori, o paesi interi, ha costi altissimi in termini dei danni che può recare al sistema scientifico e pochissimi, incerti, benefici»[41].

[39] S. Sacchi, *Comunicazione scientifica e open access. Problematiche sociali e tecnologiche nell'applicazione in Italia*, cit., p. 23.

[40] Cfr M. Amin, M. Mabe, I*mpact Factors: Use and Abuse*, «Perspectives in Publishing», 1, 2000, pp. 1-6, <http://www.elsevier.com/homepage/about/ita/editors/perspectives1.pdf>.

[41] A. Figà-Talamanca, *L'Impact Factor nella valutazione della ricerca e nello sviluppo dell'editoria scientifica*, IV seminario sistema informativo nazionale per la matematica 2000: un modello di sistema informativo nazionale per aree disciplinari, Lecce, 2000, <http://siba2.unile.it/sinm/4sinm/interventi/fig-talam.htm>.

Contare le citazioni di una rivista non può essere considerato un criterio per valutare la qualità di un articolo. Inoltre, come osserva Alessandro Figà-Talamanca

> altro è che un lavoro sia citato con una frase del tipo: "Il presente lavoro si basa sul lavoro fondamentale e pionieristico di Tizio". Altro che invece sia citato con una frase che dice: "I risultati su questo problema ottenuti da Caio non sono affidabili perché basati su un'errata metodologia". Altro ancora è che Sempronio appaia citato tra decine di altri autori come uno dei tanti ricercatori che hanno studiato lo stesso problema. Nel primo e nel secondo caso naturalmente non si può nemmeno prendere per oro colato il giudizio di chi cita il lavoro. Nel terzo caso (che è quello ormai prevalente) l'informazione fornita dalla citazione, in merito al lavoro citato, è pressoché nulla. In molti casi non si può nemmeno dire che chi cita abbia letto l'articolo citato, meno che mai che ne sia stato influenzato. Parlare di "impatto" sulla ricerca scientifica dei risultati del lavoro citato è, in questo caso, del tutto arbitrario[42].

Gli effetti negativi dell'applicazione del fattore d'impatto sono molteplici, primo tra tutti il fatto che scelte dei criteri di valutazione arbitrarie si mascherano di obiettività. Se infatti una citazione è in grado di aumentare l'IF, i ricercatori cercano di pubblicare in riviste ad alto IF citando, a questo scopo, la rivista in cui vorrebbero vedere pubblicati i propri articoli. Una pratica richiesta dagli stessi comitati di redazione delle riviste "prestigiose", e accettata dagli autori in quanto utile ad aumentare il proprio valore sul mercato accademico. Ma è possibile rendere obiettiva la valutazione? Una risposta è quella indicata da Figà-Talamanca, che scrive:

> Una valutazione discrezionale è obiettiva nella misura in cui non è dettata da interessi diretti o indiretti di chi valuta, cioè nella misura in cui si esprime un giudizio "imparziale". Nessuno è in grado a priori di assicurare questa imparzialità. Ma certamente uno degli strumenti per controllare i giudizi scientifici è quello della censura, in termini di reputazione, che può esercitare una comunità scientifica vigile e aperta. Una condizione necessaria perché questa censura possa essere esercitata è che ci sia una chiara attribuzione di responsabilità. Un giudizio basato su scelte discrezionali oscurate da presunti parametri oggettivi è sottratto ad ogni critica e quindi potenzialmente arbitrario. Queste considerazioni si applicano non solo quando si tratta di giudicare i singoli, ma anche quando il giudizio si applica a gruppi o strutture in competizione[43].

[42] *Ivi.*
[43] *Ivi.*

3. La crisi del sistema tradizionale e le trasformazioni in atto

In un lavoro recente, Whitworth e Friedman hanno definito il sistema tradizionale di pubblicazione in campo scientifico un sistema feudale perché esclusivo, datato (poco aggiornato), conservatore (in quanto resistente ai cambiamenti e all'innovazione), e i cui contenuti sono poco letti o discussi e spesso inaccessibili e molto specialistici[44]. Non si tratta di un giudizio di valore, ma di un'analisi che mette in luce alcuni limiti di tale sistema, e che merita qualche approfondimento.

Di proprietà dello *Institute for Scientific Information*, un'azienda privata che dal 1992 appartiene a quella che oggi è la Thomson Reuters Corporation, vero e proprio gigante editoriale, lo SCI pone infatti un problema di democraticità nell'accesso all'informazione.

> È importante avere ben presente questo fatto perché l'azienda, dall'alto della sua posizione dominante sul mercato, come detentrice di una formidabile base di dati sulle pubblicazioni e citazioni scientifiche, ha preso, come tutte le aziende, numerose decisioni dettate da interessi venali e basate su calcoli dei costi e dei benefici marginali.[45]

L'accentramento dell'indice ha così creato profonde distorsioni nell'organizzazione della comunicazione scientifica. Riguardando un numero limitato di riviste fondamentali, l'indice delle citazioni e il fattore d'impatto si sono rivelati efficaci strumenti di controllo capaci di attribuire molto potere a un ristretto numero di editori. L'adozione dello SCI come metro attraverso cui valutare la ricerca ha inoltre ridotto la possibilità di trovare informazioni al di fuori del nucleo fondamentale delle riviste SCI, limitando l'impatto di un sistema basato sulle associazioni; se la ricerca si nutre spesso di annotazioni secondarie e segue piste parzialmente esplorate, l'esclusione di intere aree della conoscenza umana dall'indice ha creato una "scienza di serie A" e una "scienza di serie B", rendendo l'accesso alla seconda pressoché impossibile. In altre parole, è ISI-SCI a stabilire i termini della valutazione e a fornire poi gli strumenti per analizzare le attività scientifiche. Il modello corrente di *gate-keeping* gestito da Thomson Reuter appare inoltre poco efficiente nel momento in cui il sapere si espande – e il sistema di pubblicazione accademica dovrebbe cercare di essere inclusivo e democratico piuttosto che esclusivo e plutocratico.

Se le pubblicazioni scientifiche su carta hanno rappresentato fino a tempi piuttosto recenti l'unico mezzo per diffondere l'informazione scientifica e riconoscerne la scientificità attraverso il filtro del *peer review*,

[44] B. Whitworth, R. Friedman, *Reinventing Academic Publishing online. PART I. Rigor, Relevance and Practice*, cit.

[45] A. Figà-Talamanca, *L'Impact Factor nella valutazione della ricerca e nello sviluppo dell'editoria scientifica*, cit.

alle riviste è stato in tal senso accordato un ruolo fondamentale, soprattutto nelle STM. Abbiamo osservato come il prestigio di queste ultime sia legato alla loro presenza nel catalogo SCI. Si osservi inoltre che se fino alla prima metà del secolo scorso la pubblicazione scientifica non era redditizia, l'affermazione del catalogo ISI ha trasformato questa situazione producendo un nucleo di *core journals* che le biblioteche scientifiche dovevano avere a qualsiasi costo.

Si intuisce facilmente come il catalogo, nato in origine per scopi bibliografici, abbia finito per contribuire a produrre politiche dei prezzi di tipo monopolistico tipiche di un mercato anelastico, con pesanti conseguenze: proprio a causa del loro ruolo decisivo nella valutazione del fattore d'impatto, le biblioteche universitarie non possono non abbonarsi alle riviste ISI-SCI, quale che sia il costo di tali abbonamenti. La scarsa elasticità della domanda e il carattere oligopolistico delle multinazionali dell'editoria scientifica hanno reso possibile un aumento sproporzionato dei prezzi di abbonamento delle riviste.

Gli esiti di questa situazione sono conosciuti tra i bibliotecari come "crisi dei prezzi dei periodici", espressione tramite la quale si definisce l'aumento del prezzo delle riviste che, tra 1975 e il 1995, è cresciuto del 300% oltre l'inflazione[46]. Tale fenomeno produce gravi conseguenze sia sulla politica di abbonamenti alle riviste non incluse nel catalogo SCI, sia sull'acquisto delle monografie, che restano il veicolo principale della pubblicazione umanistica, acquisto che viene sacrificato a favore degli abbonamenti alle riviste.

Un ulteriore effetto di tale politica editoriale e scientifica è stata il progressivo rafforzamento della gerarchia a due livelli tra riviste di serie A e riviste di serie B, che ha reso più duri gli effetti dei principi di inclusione/esclusione già all'opera nella creazione delle prime bibliografie.

Infine, a causa dei requisiti imposti dai grandi editori oligopolisti nel settore, tra cui oltre a Thomson Reuters Corporation è opportuno ricordare Reed Elsevier e Springer - Kluwer Academic Publishing, i *core journals* coincidono con pubblicazioni in lingua inglese. Ciò ha portato alla definizione di un sistema anglo-centrico, in cui pubblicare in lingue diverse dall'inglese costituisce di per se un fattore discriminante.

L'avvento di Internet e del Web ha in parte cambiato la situazione, sostituendo agli indici bibliometrici i cosiddetti indici webometrici, che computano le citazioni su basi di dati online[47]. Se SCI ha creato rapida-

[46] Dati OCSE. Cfr. OECD, *Report on scientific publishing of 2004-2005,* cit.

[47] L'analisi *webometrica* è nata sulla scia della bibliometria per misurare la produzione scientifica che viene disseminata tramite il Web. L'espressione "webometrics" è stata in particolare coniata nel 1997 da Almind e Ingwersen (T.C. Almind, Peter Ingwersen, *Infometric analyses on the World Wide Web: Methodological approaches to 'webometrics'*, «Journal of Documentation», 53, 4, 1997, pp. 404–426). Cfr. la voce Webometrics di wikipedia <http://en.wikipedia.org/wiki/Webometrics>. Cfr. anche

mente la sua versione digitale a pagamento (nota come Web of Science), un altro gigante dell'editoria, Reed Elsievier propone oggi Scopus, un database sviluppato quale concorrente diretto di SCI con l'obiettivo di prendere il posto attualmente occupato sul mercato dallo *Science Citation Index.* Questo obiettivo è stato perseguito attraverso due strategie principali: 1) la sostituzione dell'*Impact Factor* con lo *h-index*[48], un indice sviluppato nel 2005 da Jorge E. Hirsch che, a differenza del primo, si basa sul numero delle pubblicazioni ed il numero di citazioni ricevute da un ricercatore, permettendo così di quantificare la prolificità e l'impatto del lavoro di un singolo scienziato indipendentemente dall'etichetta della rivista; 2) una maggiore copertura delle pubblicazioni, non più confinata alle sole riviste in lingua inglese, che arriva a comprendere non solo le riviste tradizionali (oggi quelle presenti nella banca dati sono oltre 17.000), ma anche riviste *open access*[49] (circa 1200) e diversi altri tipi di prodotti editoriali tra cui atti di conferenze, pubblicazioni commerciali di diversa natura, collane di libri. Una soluzione che appare appetibile, come mostra la decisione, resa nota nel settembre del 2009, di acquisire i dati di Scopus per il Comitato di Indirizzo per la Valutazione della Ricerca (CIVR).

Tuttavia, i criteri per essere inclusi nel data base restano oscuri e discutibili. È ad esempio richiesto che la pubblicazione sulle riviste avvenga con scadenze regolari, criterio che, oltre non avere alcuna utilità sulle riviste digitali, che non sono sottoposte ai medesimi vincoli della stampa, in Italia taglia fuori i periodici pubblicati da società e associazioni scientifiche, che non abbiano un giornalista iscritto all'albo come direttore. La sfida capitanata da Reed-Elsevier non mette in discussione la divisione creata da SCI, ma si limita a cercare di volgerla a suo vantaggio, come dimostra la lista delle istituzioni che sono state coinvolte nella progettazione della base di dati, quasi solo europee e Nordamericane[50]. Certamente, la nascita di ulteriori indici come il *Chinese Science Citation Database*, che nel 2001 copriva 991 titoli, solo 31 dei quali presenti su SCI, e di indici analoghi validi per paesi e continenti cosiddetti periferici può modificare questa situazione.

Ma ad entrare nella scena sono stati anche altri attori non tradizionali, tra i quali è utile ricordare Google con il progetto Google Scholar[51].

EU TSER project PL97-1296 *The self-organization of the European Information Society* (SOEIS), 1997-99, <http://utopia.duth.gr/~mboudour/soeis/>.

48 Cfr. J.E. Hirsch, *An index to quantify an individual's scientific research output*, 2005, <http://arxiv.org/abs/physics/0508025>. È possibile calcolare l'indice h anche sull'intero Web, per esempio tramite software come *Publish or Perish*, scaricabile all'URL <http://www.harzing.com/pop.htm>.

49 Sono riviste gratuite online, compatibili con il protocollo OAI-PMH. Si veda il capitolo 4.

50 J.C. Guédon, *Open Access. Contro gli oligopoli nel sapere*, cit., pp. 77-8.

L'uso della banca dati di Google Scholar[52] è gratuito, l'inclusione delle fonti è aperta a biblioteche ed editori e intere biblioteche hanno scelto di far digitalizzare le proprie collezioni da Google; il database può essere utilizzato per produrre più indici; inoltre, gli algoritmi utilizzati per classificare le pagine dal più grande motore di ricerca presente oggi sul mercato fanno riferimento all'uso mondiale della rete e non restituiscono un'immagine del sapere scientifico confinato al solo mondo occidentale, creando un terreno di competizione scientifica più equilibrato.

Prima di analizzare nel dettaglio questa e altre possibilità oggi disponibili grazie alle nuove tecnologie, è tuttavia opportuno considerare da vicino l'architettura, i principi del Web e i suoi sviluppi più recenti. Una tale disamina permetterà di comprenderne appieno limiti e potenzialità, e di esplorare possibili modelli alternativi a quelli proposti.

[51] «Google Scholar è un motore di ricerca accessibile liberamente che tramite parole chiave specifiche consente di individuare testi della cosiddetta letteratura accademica come articoli sottoposti a revisione paritaria, tesi di laurea e dottorato, libri, preprint, sommari, recensioni e rapporti tecnici di tutti i settori della ricerca scientifica. Google Scholar consente di reperire articoli da una vasta gamma di case editrici che si rivolgono al mondo dello studio e della ricerca da associazioni scientifiche e professionali, depositi di preprint e università, oltre che nella galassia di articoli scientifici e culturali distribuiti sul Web.

Gli articoli in esso indicizzati comprendono pure quelli pubblicati da Elsevier (il più grande editore scientifico a livello globale), e nella sua funzione è simile al motore di ricerca interno Scirus di Elsevier e a Citeseer. È anche molto simile a quelli a pagamento Scopus e Thomson ISI Web of Science. Google Scholar in ogni caso afferma di coprire più siti web, sorgenti giornalistiche e lingue di tutti gli altri». Cfr. la relativa voce di Wikipedia: <http://it.wikipedia.org/wiki/Google_Scholar>.

[52] Cfr. <http://scholar.google.com> e anche la pagina <http://scholar.google.it/intl/it/scholar/about.html>.

CAPITOLO 2

UN SISTEMA UNIVERSALE DI DOCUMENTAZIONE: IL WORLD WIDE WEB

1. Un po' di storia

Il progetto di un sistema universale di documentazione risale alla fine degli anni Ottanta del Novecento. È lo stesso inventore del Web a narrare le circostanze della sua nascita e a raccontare la storia della sua evoluzione in un testo divulgativo pubblicato nel 1999 e tradotto due anni dopo in italiano[1].

Laureatosi in fisica nel 1976 al Queen's College di Oxford, il giovane scienziato inglese si impiega come programmatore alla Plessey Telecommunications prima e poi alla D.G. Nash, presso cui costruisce il suo primo computer. La sua iniziale collaborazione con il CERN risale al 1980, anno in cui scrive *Enquire*, un programma «retiforme» che anticipa una caratteristica essenziale del World Wide Web, la possibilità di collegare pezzi di informazione in forma ipertestuale. Incaricato di occuparsi dei sistemi informativi del laboratorio franco-svizzero che ospitava diverse migliaia di ricercatori delle più differenti nazionalità (i quali lavoravano su numerosi progetti collegati ma indipendenti e che utilizzavano un'ampia gamma di programmi e di calcolatori incapaci di comunicare tra loro), Berners-Lee si scontra con le difficoltà derivanti dal dover gestire una tale mole di informazione: «Preparai *Enquire* nel tempo libero, per utilizzo personale e al nobile scopo di aiutarmi a ricordare i rapporti intercorrenti tra le varie persone, calcolatori e progetti all'interno del laboratorio»[2]. E continua:

> Su *Enquire* potevo scrivere una pagina d'informazioni a proposito di una persona, una macchina o un programma. Ogni pagina era un "nodo" nel programma, una specie di scheda. L'unico modo per crearne uno nuovo era attuare un collegamento da un nodo già esi-

[1] T. Berners-Lee, *L'architettura del nuovo Web*, Feltrinelli, Milano 2001, scritto in collaborazione con Mark Fiaschetti (ed. or. 1999). Si vedano anche la documentazione disponibile all'URL <http://www.w3.org/History/> e l'ipertesto, a cura dello stesso Berners-Lee <http://www.w3.org/DesignIssues>.

[2] *Ivi*, p. 18.

Francesca Di Donato, *La scienza e la rete : l'uso pubblico della ragione nell'età del Web*,
ISBN 978-88-8453-494-1 (print) ISBN 978-88-8453-500-9 (online)

> stente. I link da e verso un nodo apparivano come un elenco numerato in fondo alla pagina, un po' come la lista delle citazioni alla fine di una pubblicazione accademica[3].

Nella costruzione del suo organigramma, il fisico inglese scopre le potenzialità insite nella possibilità di inserire collegamenti in modo libero. Non solo: al principio Berners-Lee distingue diverse tipologie di collegamenti: «Per ogni link, potevo descrivere il tipo di rapporto. Per esempio, dichiarando se la relazione tra Joe e un programma era il fatto che lo usasse o che l'avesse scritto». *Enquire* funzionava in base a un principio molto semplice: bastava etichettare ogni pezzo d'informazione con un nome e poi ordinare al computer di trovarlo. Diventando possibile fare riferimento a ogni cosa allo stesso modo, si dava così vita a una rete di informazioni. Un'altra implicazione positiva stava nel fatto che i computer avrebbero potuto seguire e analizzare gli incerti rapporti di connessione che definiscono gran parte dei meccanismi sociali, rivelando modalità di vedere il mondo del tutto inedite.

La natura e l'evoluzione del Web sono correlate in maniera inestricabile, come spiega il suo stesso autore: «L'invenzione del World Wide Web ha comportato la crescente comprensione da parte mia del potere insito nel disporre le idee in maniera reticolare, una consapevolezza che ho appunto acquisito tramite questo genere di procedura»[4]. *Enquire* lo porta a concepire una visione fondata sulla crescita decentrata e organica di idee, tecnologia e società. Alla base del sistema stava infatti l'ipotesi che i computer potessero diventare molto più potenti se posti in grado di mettere in connessione informazioni altrimenti scollegate. Potenzialmente, la ragnatela globale dell'informazione si presentava come un tutto collegato a tutto, nel quale i meccanismi sociali diventavano simili a quelli con cui funziona il cervello.

La prima collaborazione al CERN dura pochi mesi. Nell'abbandonare il laboratorio, Berners-Lee lascia al suo successore il codice di *Enquire,* che andrà in seguito perduto. Tra il 1981 e il 1984 torna a vivere in Gran Bretagna, dove lavora come ingegnere informatico scrivendo software per stampanti. È in quel frangente che acquisisce dimestichezza con i linguaggi di *mark-up,* in particolare SGML[5].

[3] *Ivi*, p. 23.

[4] *Ivi*, pp. 16-7.

[5] Un linguaggio di *mark-up* descrive i meccanismi di rappresentazione di un testo. *Mark-up* (letteralmente: marcatura) è un termine nato in ambiente tipografico per segnalare in forma standardizzata le parti del testo che necessitano di una correzione. Tale tecnica richiede una serie di convenzioni, cioè un linguaggio a marcatori di documenti.

HTML, al pari di SGML, è un linguaggio di *mark-up* di tipo descrittivo, vale a dire che la scelta del tipo di rappresentazione da applicare al testo è lasciata al software. SGML (*Standard Generalized Markup Language*) è stato il primo linguaggio di

Dall'84 riprende la collaborazione con il CERN, grazie a una borsa di studio che gli consente di lavorare al "controllo e acquisizione dati", presso il gruppo incaricato al rilevamento e all'elaborazione dei dati risultanti dagli esperimenti dei fisici delle alte energie. Lì scrive un secondo programma, *Tangle* (che significa letteralmente intrico, groviglio).

> In *Tangle*, spiega ancora, se ricorreva una certa sequenza di caratteri, il programma creava un nodo che la rappresentasse. Quando essa ricompariva, invece di ripeterla Tangle attivava un semplice rimando al nodo principale. Man mano che altre frasi venivano immagazzinate come nodi, e altri puntatori le indicavano, si formava una serie di collegamenti. Il nocciolo della questione era che l'aspetto fondamentale sono le connessioni. Non sta nelle lettere, ma in come esse si uniscono a formare parole. Non sta nelle parole, ma come si uniscono per formare frasi. Non sta nelle frasi, ma come si uniscono in un documento. Immaginai di inserire in questa maniera un'enciclopedia, quindi porre una domanda a *Tangle*. La domanda sarebbe stata ridotta in tanti nodi, che poi avrebbero fatto riferimento ai punti in cui apparivano all'interno dell'enciclopedia. Il groviglio risultante avrebbe contenuto tutte le risposte relative[6].

Il risultato dell'esperimento è fallimentare. Berners-Lee dismette il programma e tuttavia non abbandona il progetto, per il cui sviluppo l'ambiente del CERN si rivela particolarmente fecondo: la compresenza di migliaia di persone, che utilizzavano un numero imprecisabile di macchine su cui giravano i programmi più disparati e comunicavano grazie a una moltitudine di protocolli di rete, anticipava quella diversità interconnessa che di lì a pochi anni si sarebbe ritrovata anche nel mondo esterno.

> Oltre a tener conto dei rapporti tra le varie persone, esperimenti e macchine, volevo facilitare l'accesso ai vari tipi di informazione, come gli scritti tecnici dei vari ricercatori, i manuali dei diversi moduli di software, i resoconti delle riunioni, gli appunti e così via. Inoltre, mi toccava rispondere di continuo alle medesime domande poste da tante persone distinte. Sarebbe stato molto più semplice se tutti avessero potuto consultare il mio database.[7]

Sono probabilmente queste le ragioni per cui Berners-Lee persevera nella sua idea di creare un sistema di documentazione universale.

Il successore di *Tangle* è *Remote Procedure Call* (RPC), un programma scritto per facilitare la comunicazione tra i computer e le reti. Ber-

mark-up descrittivo standardizzato a livello internazionale e ha avuto ampio utilizzo nella produzione di documentazione tecnica.

[6] T. Berners-Lee, *L'architettura del nuovo Web*, cit., p. 26.

[7] *Ivi*, p. 27.

ners-Lee ha ben chiaro che costringere gli scienziati a riadattare il loro sistema di lavoro in base alle sue necessità lo porterebbe necessariamente all'insuccesso. Viceversa, deve pensare a un sistema di documentazione che consenta a ciascuno di conservare il proprio metodo organizzativo e i suoi programmi. «Dovevo creare, scrive, un sistema con regole comuni, accettabili per tutti, cioè il più possibile vicino alla mancanza di regole»[8]. RPC assolve al compito richiesto.

Il modello che sceglie per il sistema minimalista che ha in mente è l'ipertesto. Il termine, coniato nel 1965 da Ted Nelson, indica un formato nuovo, non lineare, scritto e pubblicato tramite macchine cosiddette "letterarie"[9]. Con Xanadu[10], il progetto ipertestuale di Nelson, ogni informazione sarebbe stata pubblicata in forma ipertestuale. Ogni citazione sarebbe stata dotata di un link alla fonte, garantendo agli autori originali una piccola ricompensa ogni volta che la fonte venisse letta.

Berners-Lee immagina l'ipertesto come uno strumento per il lavoro di gruppo e collaborativo, in cui ciascuno è in grado di scrivere e di leggere in un reticolo di documenti. Il sistema che ha in mente dovrebbe essere decentrato, in modo che ogni nodo sia intrinsecamente equivalente agli altri e che sia possibile collegarsi a un nodo (pagina) qualsiasi senza richiedere alcuna autorizzazione a un'autorità centrale. È su tale base che progetta un sistema in grado di combinare i link esterni con l'ipertesto e con gli schemi di interconnessione sviluppati per RPC.

> L'ipertesto sarebbe stato molto potente, aggiunge ancora, se fossimo riusciti a indirizzarlo verso il nulla più totale. Ogni nodo, documento o come preferite chiamarlo, sarebbe stato intrinsecamente equivalente agli altri. Tutti avrebbero posseduto un indirizzo di riferimento. Sarebbero esistiti insieme nel medesimo spazio, lo spazio dell'informazione[11].

Una tappa fondamentale per la definizione del suo progetto è la penetrazione del sistema operativo Unix e soprattutto di Internet all'interno del CERN. L'invenzione delle tecnologie alla base di Internet risale ai tre decenni precedenti ed è dovuta all'inventiva e all'intraprendenza di un gruppo di scienziati giovanissimi delle più prestigiose università americane (il MIT, la UCLA e la Stanford University) e di DARPA (*Defense Advanced Research Projects Agency*), l'agenzia governativa del Diparti-

[8] *Ivi*, p. 28.

[9] Si vedano: T. Nelson, *A File Structure for the Complex, The Changing and the Indeterminate*, «ACM 20th National Conference», pp. 84-100, 1965; Id, *Literary machines 90.1. Il progetto Xanadu*, Muzzio, Padova 1992 (edd. origg. 1981, 1987, 1990 e 1993).

[10] Si veda la relativa voce di Wikipedia <http://it.wikipedia.org/wiki/Progetto_Xanadu> (10/09).

[11] T. Berners-Lee, *L'architettura del nuovo Web*, cit., pp. 28-29.

mento della Difesa degli Stati Uniti incaricata dello sviluppo di nuove tecnologie per uso militare. Oggetto di una ricerca congiunta e parallela, è il progetto di reti di calcolatori per lo scambio di documentazione e informazioni che trova spazio in una «repubblica degli informatici dipendente da contratti federali e protetta dagli influssi del mondo esterno», al cui «interno si forma l'idea secondo cui il modello di socialità sviluppatosi attorno e mediante Arpanet si può trasferire al mondo comune».[12]

Uno strumento di comunicazione essenziale per gli scienziati che fanno parte della "repubblica" sono le *Request For Comments* (RFC), una serie di note inaugurata nel 1969 come strumento attraverso il quale i ricercatori si scambiano in modo rapido e informale appunti, osservazioni, idee, che vengono discussi, ampliati, definiti e poi messi in pratica. I giovani scienziati fanno dei princìpi del libero pensiero, del metodo critico e della condivisione e cooperazione con i loro pari, una pratica organizzata di lavoro. Nate all'inizio su carta, le RFC circolano poi su ARPANET e tra le reti tramite il protocollo *File Transport Protocol* (FTP) e successivamente sul Web.

L'uso delle RFC scambiate pubblicamente è fondamentale per l'affermarsi dei protocolli, tutti rigorosamente pubblici. La definizione dei protocolli si basa infatti su una procedura fissa, fatta dei seguenti passaggi. La "richiesta di commenti" gira in rete e raccoglie le osservazioni degli altri, dei cui contributi si arricchisce. Una volta raggiunto il consenso della comunità sul suo contenuto, si produce un documento, in termini tecnici una "specifica"[13], che poi è usata e testata dai gruppi di ricerca come base su cui implementare il software e, se si afferma all'interno della comunità, diventa un protocollo. La pratica delle RFC genera un effetto benefico, sia fornendo le regole della comunicazione alla comunità, sia ampliando il numero di coloro che collaborano alle innovazioni introdotte. Nel tempo, la produzione di RFC si è andata orientando sempre più sulla produzione di standard dei protocolli ("specifiche" ufficiali). L'accesso aperto (libero da *copyright* e gratuito) alle RFC favorisce la crescita di Internet perché consente alle specifiche di essere discusse, modificate e adottate sia nelle università e nei centri di ricerca, sia nelle imprese.

Ma l'elemento determinante per la nascita di Internet, forse il pilastro fondamentale della sua architettura, è lo *stack* (letteralmente "pila") di protocolli di comunicazione che si chiamerà *Transmission Control Proto-*

[12] A. Mattelart, *Storia della società dell'informazione*, Einaudi, Torino 2002 (ed. orig. 2001), p. 51.

[13] Col termine "specifica" in ingegneria del software si intende un accordo tra un produttore di servizi ed un utente. A seconda del contesto, si parla di specifiche dei requisiti (tra sviluppatore e committente), specifiche di progetto (tra progettista e sviluppatore, come nel caso delle specifiche dei protocolli), e di specifiche di modulo (tra i programmatori che hanno prodotto il modulo e quelli che lo integrano). Cfr. la relativa voce di Wikipedia: <http://it.wikipedia.org/wiki/Specifica>.

col/Internet Protocol (TCP/IP) dal 1978, progettato a partire dal 1972 da Robert Kahn, anch'egli a DARPA. Kahn dapprima introduce in un progetto di rete radio l'idea di *"open-architecture networking"*, e crea un primo programma di nome *"Internetting";* di esso, lo scienziato decide poi di sviluppare una nuova versione che soddisfi le richieste di una rete ad architettura aperta, vale a dire:

(1) che rispetti l'autonomia tecnica di ogni singola sotto-rete, senza comportare il bisogno di apportare modifiche per connettersi tramite il protocollo;

(2) in cui la comunicazione avvenga sulla base della regola del *best effort* (migliore prestazione): un pacchetto perso viene ritrasmesso;

(3) in cui per connettersi alla rete siano usati *black boxes* (*gateway*[14] e *router*[15]). I *gateway* non conservano informazioni sul passaggio dei pacchetti;

(4) che non preveda alcun controllo globale o centrale sulle operazioni.

In pratica, il TCP viene progettato con le seguenti funzionalità:

> a) Piccole sotto-sezioni con cui l'intera rete sarebbe stata capace di comunicare attraverso computer specializzati il cui compito era solo di indirizzare i pacchetti (*gateway*, *router, switch,...*).
> b) Nessuna parte della rete sarebbe stata singolo punto di rottura o sarebbe stata capace di controllare tutta la rete.
> c) Ogni pezzo dell'informazione inviata attraverso la rete avrebbe avuto una sequenza di numeri per assicurare che sarebbero stati inviati nell'ordine giusto alla destinazione e riconoscere la loro perdita.
> d) Un computer che inviava un'informazione ad un altro avrebbe saputo se questa fosse stata ricevuta nel momento in cui il computer di destinazione avesse inviato uno speciale pacchetto detto *acknowledgement*.
> e) Se l'informazione inviata da un altro computer andava persa, l'informazione veniva ritrasmessa dopo che la perdita veniva intercettata da un timeout che avrebbe identificato che l'atteso *acknowledgement* non era stato ricevuto.
> f) Ogni pezzo dell'informazione inviata attraverso la rete sarebbe stato accompagnato da un *checksum* [somma di controllo], calcolato dall'originale computer sorgente e verificato dalla destinazione, per assicurare che non fosse rovinato.[16]

[14] Dispositivi di rete il cui scopo principale è quello di veicolare i pacchetti di rete all'esterno della rete locale (LAN). Cfr. Wikipedia all'URL <http://it.wikipedia.org/wiki/Gateway> (10/09).

[15] Letteralmente "instradatore", è un dispositivo di rete che si occupa di instradare pacchetti. Cfr. Wikipedia: <http://it.wikipedia.org/wiki/Router> (10/09).

[16] Cit. da Wikipedia, voce italiana di Robert Kahn: <http://it.wikipedia.org/wiki/Robert_Kahn> (10/09).

Una volta definiti i principi[17], nella primavera del 1973 è Kahn a chiedere a Vinton Cerf (al tempo alla Stanford University) di aiutarlo a scrivere l'implementazione dello *stack* di protocolli TCP/IP[18], che viene adottato da ARPANET il primo gennaio 1983 e che diventa, nell'arco di circa dieci anni, il protocollo standard di comunicazione tra reti di computer, vale a dire l'insieme delle convenzioni tramite cui i computer si inviano i dati.

L'ultima tappa importante nella storia di Internet che ricordiamo qui è l'introduzione, sempre nel 1983, del *Domain Name System* (DNS), un sistema che consente di adottare un meccanismo scalabile e distribuito per tradurre gerarchie di nomi di *host* in un indirizzo internet. Il DNS consente di associare numeri (indirizzi degli *host*) a nomi di dominio. A ogni dominio di nomi viene associata un'autorità di dominio. È questo l'unico elemento dell'architettura di Internet che implica una forma di centralizzazione e di accentramento.

Le agenzie federali, negli anni a seguire, danno vita a politiche decisive per la diffusione e lo sviluppo di Internet, dividendo e accollandosi i costi delle infrastrutture comuni come i circuiti transoceanici, attività per coordinare la quale nasce il *Federal Networking Council* (FNC), che coopera con RARE in Europa. Gli stessi inventori di Internet scrivono due report importanti per il suo futuro sviluppo: *Toward a National Research Network*[19] (1988) commissionato dalla *National Science Foundation*, rapporto essenziale nell'influenzare Al Gore nel progetto delle autostrade informatiche e, nel 1994, *Realizing the Information Future: The Internet and Beyond*[20].

Nel 1995 il FNC emana all'unanimità la seguente definizione:

> "Internet" refers to the global information system that -- (i) is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions/follow-ons; (ii) is able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions/follow-ons, and/or other IP-compatible protocols; and (iii) provides, uses or makes accessible, either publicly or privately, high level services layered on the communications and related infrastructure described herein[21].

[17] Cfr. R.E. Kahn, *Communications Principles for Operating Systems*, «Internal BBN memorandum», 1972.

[18] Cfr. V.G. Cerf and R.E. Kahn, *A protocol for packet network interconnection*, «IEEE Trans. Comm. Tech.», COM-22, 5, 1974, pp. 627-641. DARPA fa tre contratti a Stanford (Cerf), BBN e UCL per implementare TCP/IP, dai quali hanno inizio una sperimentazione e uno sviluppo di lungo periodo.

[19] Online all'URL <http://www.nap.edu/openbook.php?isbn=NI000393>.

[20] Online all'URL <http://www.nap.edu/openbook.php?record_id=4755>.

Si osservi che in tale definizione l'unica caratteristica tecnica menzionata è la suite di protocolli (TCP/IP). In essa è inoltre essenziale il concetto di strati, sia tra i protocolli, sia di servizi di alto livello. A partire dallo *stack* di protocolli TCP/IP, Internet si fonda dunque sul concetto di architettura multi-livello, come si vede nelle immagini che seguono.

v · d · m	Suite di protocolli Internet
Livello applicazioni	DHCP, HTTP, HTTPS , SMTP, POP3, IMAP, FTP, SFTP, DNS, SSH, IRC, SNMP, SIP, RTSP, Rsync, Telnet, HSRP, RTP, BGP, RIP, IGRP, VoIP,...
Livello di trasporto	TCP, UDP, SCTP, DCCP ...
Livello di internetworking	IPv4, IPv6, ICMP, ICMPv6, IGMP, IPsec...
Livello di collegamento	Ethernet, WiFi, PPP, Token ring, ARP, ATM, FDDI, LLC, SLIP, WiMAX, HSDPA, OSPF, MPLS ...
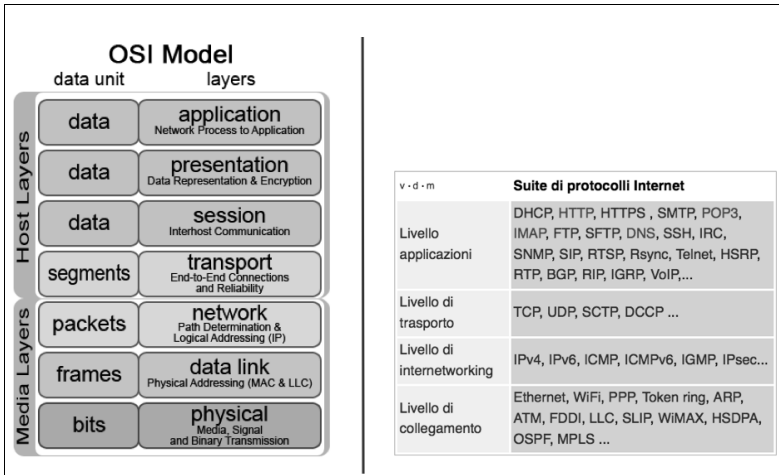
Figura 1. Riproduzione del modello Operating System Infrastructure (standard stabilito nel 1978 dall'International Organization for Standardization che stabilisce una pila di protocolli in 7 livelli) e "pila" dei protocolli TCP/IP.

Il concetto di architettura a strati (*layered network*) è essenziale in un'architettura aperta in quanto consente di introdurre innovazioni senza dover apportare cambiamenti nell'architettura della rete. Il Web, ad esempio, si trova al livello delle applicazioni (vedi figura sopra, a destra, il protocollo HTTP), e ha potuto diffondersi sopra e grazie a Internet senza richiedere alcuna modifica alla rete sottostante.

L'architettura di Internet dunque è decentrata, distribuita e prevede una sovrabbondanza di funzioni nel *network* per minimizzare il rischio di sconnessione. Essa è inoltre aperta in quanto:

1) TCP/IP fissa requisiti minimi per il collegamento alla rete, fornendo gli standard compatibili per differenti sistemi di connessione alla rete. La flessibilità di TCP/IP permette di integrare differenti sistemi di connessione alla rete e accogliere altri protocolli (ragione per la quale alla fine si sono imposti come standard comuni).

[21] Cit. in B.M. Leiner, V.G. Cerf, D.D. Clark, R.E. Kahn, L. Kleinrock, D.C. Lynch, J. Postel, L.G. Roberts, S. Wolff, *A Brief History of the Internet*, cit., in particolare nella sezione *History of the Future*.

2) Lo *stack* di protocolli è di dominio pubblico, vale a dire che i protocolli sono liberi da *copyright*. È questa una precondizione essenziale allo sviluppo della rete, poiché i protocolli e la loro documentazione tecnica hanno potuto essere consultati, discussi, migliorati e implementati, e perché altri sviluppatori hanno potuto scrivere protocolli e programmi per costruire servizi e applicazioni sopra di esse.

Ogni modifica viene comunicata alla rete, e gli utenti di Internet ne sono, dal principio in avanti, i produttori chiave. Senza tale condizione, i 15 nodi di ARPANET del 1971 non sarebbero diventati oggi centinaia di milioni[22].

Nata in ambiente scientifico per lo scambio di documenti digitali, Internet si sviluppa senza frontiere in un processo in cui i governi hanno poca parte. Gli indirizzi non hanno una localizzazione fisica fissa, i costi per l'utente non dipendono dal contenuto trasmesso, non esiste un metodo per identificare gli utenti, né un controllo centrale della rete (il solo coordinamento centralizzato è, lo si è accennato, il sistema di indirizzamento).

Si può dunque riassumere che le caratteristiche fondanti di Internet sono due:

a) in primo luogo, l'architettura della connessione in rete è aperta, decentrata, distribuita e multi-direzionale;

b) in secondo luogo, tutti i protocolli di comunicazione e le loro implementazioni sono aperti, distribuiti e suscettibili di modifica (anche se non tutti sono modificabili)[23].

L'infrastruttura fisica e lo *stack* di protocolli TCP/IP costituiscono le precondizioni per la nascita di ulteriori applicazioni che si servono di protocolli di livello superiore, tra i quali ricordiamo IMAP e POP per l'e-mail (che nasce nel 1972), il protocollo UUCP di Unix che consente la nascita di Usenet (1985), e molti altri ancora.

Tuttavia, alla fine degli anni 80 l'interesse verso Internet in Europa era scarso; le istituzioni scientifiche e politiche europee stavano infatti tentando di progettare, con la poca lungimiranza di un eccessivo campanilismo, una propria rete alternativa. Ma l'esistenza di protocolli standardizzati è per Berners-Lee l'occasione per dimostrare la possibilità di creare un ponte tra la pluralità di sistemi operativi e di reti esistenti.

La prima proposta ufficiale di un sistema informativo ipertestuale che il giovane collaboratore inoltra al CERN risale al marzo 1989; nonostante non riceva alcuna risposta, nel maggio 1990 Berners-Lee presenta una

[22] Per i dati, si veda questo grafico: <http://www.totale.it/wp-content/uploads/dimensioni-di-internet.jpg>. Si noti che esistono eccezioni a questo principio. Alcuni protocolli di *routing* dello strato della rete e alcuni protocolli dello strato applicativo (per esempio Skype) sono proprietari.

[23] M. Castells, *Galassia Internet*, Feltrinelli, Milano 2006 (ed. orig. 2001), p. 39.

seconda proposta[24], anch'essa ignorata. Tuttavia, ottiene l'appoggio informale del capo della sua divisione, che gli permette di acquistare una macchina NeXT sulla quale comincia a programmare il software del suo progetto, che chiama World Wide Web. Con il sostegno pratico di Robert Cailliau, un ingegnere impiegato al CERN particolarmente sensibile al problema dell'interoperabilità tra hardware e software che si convince immediatamente della bontà dell'idea, decide di cercare un appoggio esterno al laboratorio di fisica, di nuovo senza successo.

Al 1990 risale la definizione dei tre princìpi e pilastri del Web: lo schema per definire gli indirizzi dei documenti (*Universal Resource Identifier* o URI[25]), il protocollo di trasmissione dei dati (*Hypertext Transfer Protocol* o HTTP) e il linguaggio di contrassegno che definisce la formattazione delle pagine contenenti link ipertestuali (*Hypertext Mark*-up *Language* o HTML):

> Il segreto stava nel definire poche regole base, comuni, di 'protocollo', per permettere ai computer di dialogare tra di loro, di modo che quando tutti i computer di ogni luogo avessero fatto altrettanto, il sistema si sarebbe arricchito, invece di collassare. Per il Web tali elementi erano, in ordine decrescente di importanza, URI, HTTP e HTML, cioè gli identificatori universali, il protocollo di trasferimento e il linguaggio[26].

Poiché i tentativi di trovare collaboratori disposti a sviluppare un browser, programma essenziale a rendere l'utilizzo del Web effettivo, naufragano, Berners-Lee decide di scriverlo da solo. Lo stesso accade per il primo server Web, il programma che conserva le pagine in una parte del computer accessibile dall'esterno.

Finalmente, registra il suo computer col nome info.cern.ch, mettendo a disposizione i suoi appunti, le specifiche di URI, HTTP e HTML e le informazioni sul progetto in corso[27]. Per dimostrarne gli usi possibili all'interno del CERN, si serve dell'elenco telefonico del laboratorio, che diviene così accessibile a tutti.

> Quel che avevamo ottenuto fino a quel momento si basava su pochi princìpi chiave appresi con l'esperienza. Il punto cruciale era l'idea di

[24] La proposta originale è disponibile sul Web all'URL <http://www.w3.org/History/1989/proposal.html>.

[25] Contrariamente alla volontà dello stesso inventore del Web, l'URI sarà rinominato *Unique Resource Locator* (URL). Sul dibattito in seno al W3C su questo punto, che ha una considerevole rilevanza filosofica, si veda: T. Berners-Lee, *L'architettura del nuovo Web*, cit., pp. 63-65.

[26] T. Berners-Lee, *L'architettura del nuovo Web*, cit., p. 44.

[27] Tutta la documentazione è accessibile nella sezione *Design Issues* sul sito del World Wide Web Consortium (W3C), all'URL <http://www.w3.org/DesignIssues>.

> universalità, la rivelazione che un solo spazio dell'informazione potesse inglobare tutto, regalandoci un potere e una coerenza inauditi. Da qui derivavano molte decisioni tecniche. [...] Era una rivoluzione copernicana rispetto alla filosofia dei precedenti sistemi informatici. La gente era abituata ad andare a cercare le informazioni, ma di rado faceva riferimento ad altri computer, e anche in tal caso doveva ripetere una lunga sequenza complessa di istruzioni per accedervi[28].

Viceversa, il nuovo sistema di documentazione avrebbe permesso di trovare l'informazione sfruttando i meccanismi mentali che tutti usiamo naturalmente per ricordare, basandosi sull'uso delle citazioni, una modalità di lavoro consueta agli accademici e, più in generale, agli scienziati, che come abbiamo visto è essenziale anche per la valutazione della ricerca. Ma in un modo radicalmente nuovo.

> La comunità dei ricercatori usa da sempre dei collegamenti del genere tra documenti cartacei: tavole dei contenuti, indici analitici, bibliografie e sezioni di consultazione e rimandi sono tutti quanti link ipertestuali. Però, sul web, i link ipertestuali possono essere seguiti in pochi secondi, invece che in settimane di telefonate e inoltro della posta. E d'un tratto gli scienziati possono sottrarsi all'organizzazione sequenziale di ogni pubblicazione e bibliografia, scegliendosi un percorso di riferimenti che faccia al caso loro[29].

2. Dal Mundaneum al Memex: alcune premesse teoriche

All'importanza del collegamento nei sistemi di documentazione e all'idea di ipertesto si erano già dedicati, prima di Ted Nelson e di Berners-Lee, scienziati di diverse discipline. Tra la fine dell'Ottocento e la prima metà del Novecento sono in particolare documentaristi e ingegneri ad affrontare la questione della catalogazione e dell'accesso all'informazione. Oltre al già citato Eugene Garfield, tra questi vale la pena ricordare in particolare Paul Otlet e Vannevar Bush.

Nato a Bruxelles nel 1868 da un'agiata famiglia di avvocati, Paul Otlet intraprende nella prima giovinezza studi giuridici. Tuttavia, egli mostra da subito una vocazione agli studi speculativi, vocazione che viene incoraggiata da Edmond Picard e Paul Héger, la cui frequentazione nell'ambiente familiare lo spingerà ad abbracciare la filosofia positivista. Nel 1891 un gruppo di eruditi belgi fonda a Bruxelles la *Société des Etudes Sociales et Politiques*, che il giovane Otlet vede nascere con grande interesse. All'interno della *Société*, Henri La Fontaine dirige la Sezione di Bi-

[28] T. Berners-Lee, *L'architettura del nuovo Web*, cit., p. 42 e pp. 44-45.
[29] *Ivi*, p. 45.

bliografia. L'amicizia e la collaborazione con La Fontaine, premio Nobel per la pace nel 1913, ne incoraggiano la creazione di repertori bibliografici in materia giuridica[30] e, più in generale, la formulazione di un progetto innovativo che si propone la costituzione di centri di documentazione capaci di mettere a disposizione dell'umanità la grande varietà di documenti prodotti nelle diverse discipline. Il suo primo saggio di bibliografia (*Un Peu de Bibliographie*), in cui esprime apprezzamento per una bibliografia di matrice positivista e per il positivismo in generale, risale al 1892.

Convinto sostenitore della necessità che le scienze umane e sociali imitino le cosiddette scienze dure, occupandosi del problema del registro delle conoscenze umane, Otlet si occuperà in seguito della catalogazione delle fonti sociologiche. Con la creazione del catalogo, Otlet si persuade della necessità di una classificazione, cioè di un vocabolario controllato che consenta di descrivere con esattezza i diversi materiali.

Nel 1893, la Sezione di Bibliografia si trasforma nell'Istituto Internazionale di Bibliografia Sociologica (IIBS). La bibliografia viene intesa da Otlet e La Fontaine come una formula sintetica che permette di rimediare ai grandi mali che affliggono le scienze sociali e, più in generale, l'umanità. Entrambi coltivano infatti parallelamente due grandi passioni: la documentazione e il pacifismo[31].

È su tali premesse che la creazione di una disciplina che permetta una comunicazione di tipo universale si colloca come parte fondamentale del lavoro quotidiano degli uomini di scienza. La scienza della documentazione, come enunciata da Otlet, nasce dunque allo scopo di studiare i problemi della ricerca e del suo avanzamento, nel piano generale della comunicazione scientifica applicata al progresso umano. A tal fine si rendono necessarie una scienza e una tecnica della documentazione.

Paul Otlet fonda così la teoria della documentazione come scienza che si occupa della classificazione dei dati relativi alla produzione, alla conservazione, alla circolazione e all'uso (o impatto) degli scritti e documenti di ogni specie. Il fine ultimo della sua impresa è unire l'umanità attraverso l'opera scientifica.

[30] Nel 1891, Paul Otlet pubblica, assieme a Pierre Blanchemerle, Joseph Cassiers e Max Hallet, *Le Sommaire périodique des revues de droit* (Il Sommario periodico delle riviste di diritto), tavole mensili di tutti gli articoli giuridici pubblicati nei mensili belgi. Per una bibliografia completa dell'opera di Otlet, a cura di W. Boyd Rayward, si veda <https://www.ideals.uiuc.edu/html/2142/652/Paul%20Otlet%20Bibiliography.-htm>.

[31] La sintesi di tale duplice interesse viene pubblicata in due saggi: il *Traité de documentation. Le livre sur le livre. Théorie et pratique*, del 1934 e *Monde: Essai d'universalisme* dell'anno successivo. Il primo è disponibile online all'indirizzo <https://archive.ugent.be/handle/1854/5612>, il secondo all'URL <https://archive.ugent.be/handle/1854/8321>.

Otlet descrive la nascita di una nuova scienza che si scompone in un processo triadico: 1. la descrizione dei libri (biblioteconomia); 2. le leggi della loro produzione, conservazione e diffusione (bibliografia); 3. le applicazioni pratiche delle conoscenze bibliografiche (documentazione). Le fasi del processo si propongono di risolvere altrettanti problemi metodologici che riguardano le scienze bibliografiche: il registro, la moltiplicazione e la disponibilità dell'informazione.

La scienza della documentazione si configura come disciplina che si occupa dello studio del libro e del documento e che assume come sue parti biblioteconomia e bibliografia, che il documentarista teorizza e sistematizza nel *Traité de documentation,* il primo trattato scientifico di teoria della documentazione pubblicato a Bruxelles nel 1934.

Con una metafora efficace, Otlet definisce la sua nuova disciplina «metallurgia documentale»: un processo di estrazione, separazione e trattamento dei materiali intellettuali grezzi, che avviene tramite l'elaborazione e l'interscambio di conoscenze. Così, la documentazione viene intesa come corpo epistemologico che è al contempo oggetto di studio e strumento metodologico.

Su un piano applicativo, Otlet e La Fontaine danno vita a un insieme di creazioni innovative destinate a orientare la scienza documentale dei decenni successivi: l'Istituto internazionale di bibliografia, centro dell'attività bibliografica europea; il Repertorio bibliografico universale, un catalogo di cataloghi che già nel 1897 contiene un milione e mezzo di registri bibliografici; la scheda bibliografica, un innovativo strumento documentale che ancora oggi costituisce uno standard della biblioteconomia; la Classificazione Decimale Universale, un adattamento della classificazione universale di Dewey di cui Otlet viene a conoscenza nel 1894. Elementi, questi, che danno origine al paradigma documentale otletiano.

Ma l'esperienza più importante e significativa dell'opera di Otlet è il Palais Mundial, che sfocia successivamente nel Mundaneum, una sorta di *universitas* universale, un tempio del sapere consacrato alla conoscenza scientifica. Il Mundaneum, che avrebbe ospitato una Biblioteca Internazionale, l'istituto Internazionale di Bibliografia, il Repertorio Bibliografico Universale, una Enciclopedia di Documentazione, l'officina di Associazioni Internazionali e l'Università Internazionale, si configura come un luogo animato dai princìpi di totalità, simultaneità, gratuità, volontarietà, universalità, e da un ideale di vita internazionale, sostenuta dalla ragione e dedicata alla cooperazione e alla pace nel mondo. Un tempio consacrato all'intelletto, nel quale i ricercatori e gli intellettuali del mondo potessero avere la possibilità di studiare tutto quanto prodotto e pensato dagli uomini nel corso dei secoli.

Il progetto ottiene grandi riconoscimenti internazionali, ma uno scarso appoggio istituzionale. Otlet getta tuttavia le basi del lavoro documentale, indicando gli attrezzi del mestiere necessari al lavoro di ricerca. Diversi studi su Otlet lo hanno indicato come precursore dell'idea di iperte-

sto[32]. L'uso delle schede e dei cataloghi proposto dal padre della scienza documentale dà vita a un nuovo paradigma, in cui la scheda bibliografica corrisponde al nodo e i link che creano un sistema di navigazione ipertestuale sono dati dai cataloghi. Il sistema di classificazione proposto da Otlet si distingue come un nuovo dispositivo di accesso all'informazione e al documento, da molti considerato come l'inizio della rottura con il modello della struttura lineare. Il suo interesse si concentra sugli aspetti di un problema centrale per la società, le cui conseguenze sono tanto tecnico-organizzative quanto teoretiche: il problema di rendere la conoscenza facilmente accessibile e consultabile dall'umanità.

È questo un tema che lo avvicina a Vannevar Bush, che nel 1945 pubblica il famoso saggio *As we may think*[33]. «La conoscenza, scrive, per essere utile alla scienza deve poter essere continuamente ampliata, in primo luogo archiviata e soprattutto consultata». Preside della facoltà di ingegneria al MIT tra il 1932 e il 1938, in seguito Bush diventa direttore capo dell'Ufficio americano per la Ricerca e lo Sviluppo scientifico e consigliere scientifico del Presidente. Nello stesso anno in cui pubblica *As we may think*, indirizza al presidente Truman un rapporto[34] in cui, con sguardo lungimirante, sostiene l'importanza della ricerca di base; un documento che nei decenni successivi si rivelerà determinante nell'orientare la politica americana in materia di innovazione e sviluppo. Il rapporto

[32] Su questo argomento si vedano: W. Boyd Rayward, *Visions of Xanadu: Paul Otlet (1868-1944) and Hypertext*, «Journal of the American Society of Information Science», 45, 1994, pp. 235-250, <http://people.lis.uiuc.edu/~wrayward/otlet/xanadu.htm>; P. Arnau Reved, *Trascendencia de la vida y obra de Paul Otlet*, <http://www.ucm.es/BUCM/revistas/byd/11321873/articulos/RGID9595220153A.PDF>; R. Day, *Paul Otlet's book and the writing of social space*, «Journal of the American Society of Information Science», 48, 4, 1997, pp. 310-17; I. Rieusset-Lemarié, *P. Otlet's Mundaneum and the International Perspective in the History of Documentation and Information Science*, «Journal of the American Society of Information Science», 48, 4, 1997 pp. 301-310; M. Buckland, *What is a Document?*, «Journal of the American Society of Information Science», 48, 9, 1997, pp. 804-809, <http://people.ischool.berkeley.edu/~buckland/whatdoc.html>; C. van den Heuvel, *Mundaneum. Architectures of Global Knowledge: The Mundaneum and the World Wide Web*, «Destination Library», 15, 2008, pp. 48-53, <http://www.virtualknowledgestudio.nl/staff/charles-van-den-heuvel/vdheuvel-mundaneum.pdf>; J. M. Zurita Sànchez, *El paradigma otletiano como base de un modelo para la organizaciòn y difusiòn del conocimiento cientìfico*, Tesi di laurea in biblioteconomia, Universidad nacional autònoma de México, 2001, <http://etdindividuals.dlib.vt.edu:9090/358/1/paradigma_otletiano.pdf>.

[33] V. Bush, *As we may think*, «The Atlantic Monthly», 176, 1, 1945, pp. 101-108, <http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush.shtml>, tr. it. *Come possiamo pensare*, in T. Nelson, *Literary machines 90.1*, cit., pp. 1/38-53.

[34] V. Bush, *Science: The Endless Frontier. A Report to the President*, Washington, U.S. Government Printing Office, 1945, <http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm>. L'impatto del report di Bush è notevolmente incisivo. Se nel 1930 il finanziamento statale alla ricerca pubblica e privata è del 14%, nel 1947 passa al 56%. cfr. A. Mattelart, *Storia della società dell'informazione*, cit., p. 46.

di Bush segna «un autentico punto di svolta nella storia, evolutiva, della Repubblica della Scienza». Indicando nella scienza di base «le fondamenta su cui edificare la sicurezza nazionale (economica e sociale, prima ancora che militare) della potenza che si accinge a vincere definitivamente la seconda guerra mondiale e a progettare il nuovo ordine mondiale per l'era di pace che sta iniziando», esso «propone un rapporto affatto nuovo tra scienza e politica». Bush sostiene la necessità che il governo federale smantelli l'*Office of Scientific Research and Development*, da lui diretto fino a quel momento, preservandone però lo spirito, e fornendo agli scienziati accademici

> risorse finanziarie e umane, perché possano realizzare in piena libertà le loro ricerche, anche quando si tratta di studi, come la matematica superiore o l'astrofisica, che non hanno un'immediata ricaduta pratica. [...] Le sue lucide e precise indicazioni saranno seguite da Harry Truman e da tutti i presidenti che lo seguiranno. La scienza assume davvero un ruolo strategico a tutto campo negli Stati Uniti. In ambito militare, certo. Ma anche e soprattutto in ambito civile. Nascono nuovi enti, come la National Science Foundation (Nsf), incaricata dal governo federale di sostenere e coordinare le attività di ricerca nelle università americane; o come i National Institutes of Health (Nih), incaricati di sostenere e coordinare a ogni livello le attività di ricerca a carattere biomedico. In definitiva, grazie alla precisa scelta politica consigliata da Vannevar Bush, la conoscenza scientifica diventa la leva dello sviluppo degli Stati Uniti d'America[35].

Ed è esattamente in quel clima e in quegli stessi organismi che vede la nascita Internet. Ma torniamo a *As we may think*. Se durante la seconda guerra mondiale Bush aveva messo la propria opera al servizio dello Stato e, al pari di altri scienziati tra cui vale la pena ricordare il matematico inglese Alan Turing, padre del moderno calcolatore e artefice della decriptazione del codice tedesco *Enigma*, aveva ricoperto un importante ruolo nella ricerca militare, *As we may think* è una riflessione sul rapporto tra la tecnologia e il problema, particolarmente attuale in epoca postbellica, della pace nel mondo. Come può la tecnica, si domanda l'ingegnere americano, contribuire al benessere dell'umanità? Bush risponde all'interrogativo sciogliendo, nel corso del saggio, un'altra questione che sta, a suo parere, a monte del problema: la questione riguarda il modo in cui il bagaglio di conoscenze umane possa creare le condizioni per la pace nel mondo. Perché le tecnologie possano mettere l'uomo in grado di costruire condizioni di vita migliori, una «casa ben arredata» in cui «vi-

[35] P. Greco, *Il modello Venezia. La comunicazione nell'era post-accademica della scienza*, in N. Pitrelli, G. Sturloni (a cura di), *La comunicazione della scienza, Atti del I e II convegno nazionale*, Zadig, 2004, <http://ics.sissa.it/conferences/csIntroduzione.pdf>, pp. 9-10.

vere in buona salute»[36] suggerisce Bush con un'immagine dall'assonanza kantiana[37], è necessaria un'indagine sul metodo della ricerca e sul processo di costruzione della scienza. Il saggio non affronta dunque un tema meramente tecnico: viceversa, l'argomentazione su cui è imperniato è anche e soprattutto una riflessione filosofica e politica sulle modalità di produzione e di trasmissione del sapere.

Vannevar Bush fonda il proprio ragionamento su una importante premessa condivisa da buona parte della tradizione filosofica occidentale da Platone in poi[38], vale a dire il fatto che la conoscenza umana è un insieme collegato in un tutto che, in quanto tale, ha una dimensione universale la quale trascende la singola vita individuale. Il sapere è tale in quanto frutto di un processo cumulativo, e si struttura, grazie alla collaborazione degli scienziati, in sistemi collegati che includono l'intero patrimonio delle conoscenze umane. Per questo, l'accesso all'informazione scientifica è una condizione necessaria alla stessa possibilità della scienza.

Se la diffusione della stampa ha permesso all'umanità di produrre enormi quantità di informazioni a un ritmo sempre crescente, l'accrescersi progressivo dei dati e delle conoscenze raccolti in libri, articoli e testi di varia natura non implica la possibilità di condividere tale patrimonio; inoltre, la crescente specializzazione del sapere, prosegue Bush, confligge con l'esigenza di sistemi di comunicazione veloci ed efficienti.

Il principale ostacolo all'accesso risiede negli alti costi di riproduzione, ostacolo che può essere tuttavia facilmente superato con l'ausilio di tecnologie in grado di comprimere grandi quantità di libri in dispositivi di piccole dimensioni. La proposta di Bush individua nell'uso delle macchine una soluzione al problema dell'archiviazione; ma il problema della consultazione del sapere è assai più serio e di difficile soluzione. Esso infatti coinvolge l'intero processo grazie al quale l'uomo trae vantaggio dal suo bagaglio di conoscenze: la selezione. Bush riconosce che il cuore del problema è più profondo del semplice ritardo nella meccanizzazione delle biblioteche. Il processo di selezione è «come uno scalpello nelle mani di un ebanista», che dev'essere pertanto considerato e affrontato con riguardo al metodo della ricerca scientifica, cui l'ingegnere americano dedica ampio spazio nella parte centrale del saggio.

Il cuore del problema deve essere ricondotto all'artificiosità dei sistemi di indicizzazione in uso negli archivi, che organizzano l'informazione gerarchicamente, in base a una struttura ad albero. Poiché le informazio-

36 V. Bush, *Come possiamo pensare*, cit., p. 1/53.

37 Nell'*Architettonica della ragion pura* Kant paragona l'edificio della conoscenza scientifica a una casa d'abitazione (I. Kant, *Critica della ragion pura*, Utet, Torino 1967, A 707- B 735).

38 Si veda in particolare il mito dell'*anamnesis* nel *Menone,* 81c-d (Platone, *Menone,* Laterza, Roma-Bari, 2004, p. 34).

ni si trovano in un unico punto dell'archivio, a meno di duplicati, si devono avere regole per decidere quale cammino ci porterà a ciò che cerchiamo. Il limite di tale struttura è che a un documento corrisponde un unico percorso.

> La mente umana non funziona in questo modo. Essa opera per associazioni. Una volta che essa abbia un elemento a disposizione, salta istantaneamente all'elemento successivo suggerito, in base a un intreccio di piste registrate nelle cellule del cervello, dalla associazione dei pensieri.[39]

Bush suggerisce così di ripensare la selezione dell'informazione a partire dal modo in cui funziona la nostra mente, e indica nel *Memex*[40], una macchina fotoelettronica in grado di attuare e seguire riferimenti incrociati su microfilm servendosi di codice binario, fotocellule e fotografia istantanea, un esempio di meccanizzazione della selezione per associazione piuttosto che per indicizzazione. L'accesso virtualmente universale al patrimonio culturale dell'umanità e l'impiego delle macchine nella ricerca hanno, ai suoi occhi, un potenziale creativo dirompente. Se infatti non esistono sostituti meccanici per il pensiero creativo, ma pensiero creativo e pensiero essenzialmente ripetitivo sono molto diversi; tuttavia ogni qualvolta si combinano e si registrano fatti sulla base di processi logici stabiliti, l'aspetto creativo «riguarda solo la scelta dei dati e del processo da impiegare, e da quel momento in poi la manipolazione è di natura ripetitiva, e dunque adatta per essere affidata alle macchine»[41].

Il *Memex* è un dispositivo meccanico, una scrivania meccanizzata formata da schermi translucidi inclinati, una tastiera, e gruppi di bottoni e di leve. Al suo interno sono archiviati gli strumenti di lavoro del ricercatore (i libri e gli articoli su cui lavora, come gli appunti che produce), proiettabili velocemente su schermi inclinati. Grazie al *Memex* è possibile consultare il proprio archivio tramite lo schema di indicizzazione tradizionale: basterà digitare il codice del libro desiderato, e la sua copertina sarà rapidamente proiettata su uno dei visori. Ma la caratteristica fondamentale del "*Memory extender*" è che esso consente al ricercatore di collegare stabilmente due testi qualsiasi premendo un semplice tasto[42]. La valenza pratica della macchina ideata da Bush è descritta dall'autore come segue:

[39] V. Bush, *Come possiamo pensare*, in T. Nelson, *Literary machines 90.1*, cit., p. 1/49.

[40] Il nome deriva dalla contrazione dell'espressione "*Memory extender*".

[41] V. Bush, *Come possiamo pensare*, cit., p. 1/49.

[42] A differenza dei link sul Web, i collegamenti immaginati da Bush sono bidirezionali; vale a dire che ogni volta che un elemento viene proiettato, un altro elemento, ad esso collegato, può essere istantaneamente richiamato premendo un bottone, e visualizzato.

> Il proprietario del memex, per esempio, è interessato alle origini e alle proprietà dell'arco. In particolare, sta ricercando il motivo, nelle battaglie delle Crociate, per cui l'arco corto turco fosse superiore all'arco lungo inglese. Ha a disposizione dozzine di libri e articoli pertinenti nel suo memex. Dapprima sfoglia un'enciclopedia, trova un articolo interessante ma non approfondito, lo lascia proiettato. Poi, in un resoconto storico, trova un altro argomento pertinente, e lo congiunge al precedente. Procede in questo modo, costruendo un percorso di molti elementi. Di quando in quando inserisce un proprio commento, sia congiungendolo alla pista principale sia connettendolo come pista laterale a un elemento particolare. Quando diventa evidente che le proprietà elastiche dei materiali disponibili erano fondamentali per il rendimento dell'arco, egli devia su una pista laterale che lo porta attraverso libri di testo sull'elasticità e tavole di costanti fisiche. Inserisce una pagina di analisi scritta di suo pugno. In questo modo egli costruisce una pista della sua ricerca attraverso il labirinto del materiale che ha a disposizione[43].

Così, quando numerosi elementi sono stati congiunti per formare una pista, grazie al *Memex* possono essere esaminati l'uno dopo l'altro, velocemente o lentamente, muovendo una leva come quella usata per girare le pagine del libro. È come se gli elementi, parti di libri preesistenti, venissero raccolti separatamente e quindi rilegati nuovamente assieme, in modo da formare un nuovo libro.

Affrontando il problema della selezione dell'informazione, Bush restituisce al ricercatore un ruolo attivo e fondante.

> Nascerà la nuova professione di battitore di piste, persone che si ingegneranno a tracciare percorsi significativi attraverso l'immane mole dell'esperienza umana. L'eredità del maestro ai suoi discepoli non saranno più solo i suoi contributi alla conoscenza comune, ma l'intera impalcatura di conoscenze sulla quale essi sono stati costruiti[44].

L'attività della ricerca si traduce praticamente nell'inserimento di ogni singolo elemento in una o più piste; percorsi che restano così patrimonio dell'umanità, tanto quanto i libri, e danno vita a enciclopedie di concezione radicalmente nuova, dotate di una trama di piste associative che le attraversano. L'utilità insita nella possibilità di conservare traccia dei passaggi del processo di ricerca è esemplificata con riferimento alla ricerca sulle origini e le proprietà dell'arco, come segue:

[43] V. Bush, *Come possiamo pensare*, cit., p. 1/51.
[44] *Ivi*, p. 1/52.

> Qualche anno dopo, durante un suo colloquio con un amico, il discorso cade sui modi strani in cui le persone resistono alle innovazioni, anche se di vitale interesse. Egli dispone di un esempio, ed esattamente di come gli europei, pur riconoscendo che la loro arma aveva una gittata minore, si rifiutarono di adottare l'arco turco: dispone addirittura di una pista al riguardo. Un tocco mostra l'elenco dei codici. La pressione di alcuni tasti porta all'inizio della pista. Una leva permette di scorrerla a piacimento, fermandosi sui punti interessanti e dipartendosi in escursioni laterali. È una pista interessante, pertinente alla discussione. Così aziona un riproduttore, fotografa l'intera pista e la passa all'amico perché la inserisca nel proprio memex, congiungendola opportunamente ad altre sue piste[45].

L'articolo di Bush ha avuto un influsso fondamentale anche sulla messa a punto dello *Science Citation Index* (SCI), uno strumento che, come abbiamo visto, ha un influsso importantissimo nella storia recente della comunicazione scientifica. Ma anche per lo sviluppo del Web che, pur basandosi anch'esso sulla centralità delle citazioni, ha proposto un modello diverso e alternativo allo SCI, fondato sulle libere associazioni e sull'assenza di un'autorità centralizzata.

3. Architettura del World Wide Web e architettura dell'informazione

Nonostante alcune evidenti differenze, nel pensiero e nei progetti di Paul Otlet, Vannevar Bush ed Eugene Garfield è dunque possibile rinvenire gli elementi essenziali a caratterizzare la filosofia e i princìpi che stanno alla base del Web. Le implicazioni di un sistema di documentazione ipertestuale universale sono discusse dallo stesso Berners-Lee in una lezione tenuta in Giappone nel 2002[46].

> Il concetto del Web ha integrato molti diversi sistemi informativi disparati, creando uno spazio immaginario astratto in cui le differenze tra essi fossero ininfluenti. Il Web doveva includere ogni sorta di informazione su qualunque sistema. L'unica idea comune necessaria a mettere ciò assieme fu identificare un documento tramite lo *Universal Resource Identifier* (URI). Direttamente da ciò sono come piovute architetture (*design*) di protocollo (come HTTP) e di formati di dati (come HTML) che hanno permesso ai computer di scambiarsi informazioni stabilendo una corrispondenza tra i propri formati locali e standard in grado di offrire interoperabilità globale[47].

[45] *Ivi*, p. 1/51.

[46] T. Berners-Lee, *The World Wide Web - Past, present, future. Exploring Universality*, 2002, <http://www.w3.org/2002/04/Japan/Lecture.html>.

In termini semplificati, Berners-Lee spiega che il Web non è un programma ma un insieme di protocolli che richiede a ciascuno di assegnare un nome unico ai documenti. Si tratta di una richiesta inaggirabile in quanto necessaria al suo funzionamento, ma ampia: è infatti l'unico limite che l'architettura del Web impone alla rappresentazione e alla organizzazione dei dati.

Un esempio ci aiuta a capire come l'architettura del Web si basi su tre principi fondamentali, associati ai tre pilastri tecnici sopra ricordati: l'identificazione (che avviene tramite URI), l'interazione (possibile grazie ai protocolli, di Internet [TCI/IP] e del Web [HTTP]) e i formati (nell'esempio, XHTML, che costituisce una variante di HTML):

> Progettando un viaggio in Messico, Nadia consulta il servizio meteorologico di Oaxaca "http://weather.example.com/oaxaca" su una patinata rivista di viaggi. Nadia ha sufficiente dimestichezza col Web per capire che http://weather.example.com/oaxaca è un URI e che è verosimile che possa reperire informazioni relative al suo bisogno tramite un browser. Quando inserisce l'URI nel suo *browser*:
> 1. Il *browser* riconosce che ciò che Nadia ha scritto è un URI.
> 2. Il *browser* compie l'azione di recupero dell'informazione in accordo con la sua configurazione di comportamento per le risorse identificate tramite lo schema URI "http".
> 3. L'autorità responsabile di "weather.example.com" offre le informazioni in risposta alla richiesta.
> 4. Il *browser* interpreta la risposta, identificata come XHTML dal server, e se necessario offre ulteriori azioni di recupero per elementi grafici ivi contenuti e altri contenuti.
> 5. Il *browser* mostra l'informazione recuperata, che include link ipertestuali ad altre informazioni. Per trovarle, Nadia può seguire questi link ipertestuali[48].

Nella sua lezione, Berners-Lee insiste sul fatto che il passato, il presente e soprattutto il futuro del Web dipendono dal modo in cui viene garantita e protetta l'universalità del sistema. Sul piano tecnico, ciò è possibile postulando la separazione tra contenuto dell'informazione e forma in cui essa viene veicolata da due punti di vista principali: in primo luogo, mantenendo valida la condizione che ha permesso la nascita della ragnatela ipertestuale su Internet, vale a dire la sua indipendenza dal sostrato materiale (*hardware*) e dal modo in cui le idee sono codificate e

[47] *Ivi*. Le specifiche URI, HTTP e HTML formano cosiddetti "*bus*" che connettono i programmi che risiedono sul computer dell'utente (detto *client)* ad altri programmi installati su computer remoti (cosiddetti *server*) che forniscono le informazioni.

[48] I. Jacobs, N. Walsh (a cura di), *Architecture of the World Wide Web. Volume One*, 2004, <http://www.w3.org/TR/webarch/>, p. 6.

veicolate (*software*); in secondo luogo, definendo con maggior rigore i confini tra "documenti" e "dati", i primi dedicati alla lettura da parte degli uomini, i secondi interpretabili dalle macchine.

Nella conferenza giapponese, Berners-Lee ricorda l'esigenza che ha dato origine al World Wide Web, quella di mettere a disposizione degli ospiti del CERN uno strumento comune che permettesse loro di condividere documenti in rete. Si è osservato come l'idea di Berners-Lee si fondasse su un principio filosofico semplice e, assieme, forte: creare un sistema di condivisione dei documenti "leggero", fondato sulla sottrazione, piuttosto che sull'addizione di requisiti. Così operando, il sistema pensato dallo scienziato inglese sarebbe stato facilmente condivisibile da diversi computer, sistemi operativi e programmi applicativi, e, in quanto tale, universale, cioè aperto a tutti coloro che volessero connettervisi. Da un punto di vista pratico, l'universalità del World Wide Web è garantita dalle specifiche del W3C[49], che raccomandano:

1. L'indipendenza dall'hardware e dal software, importanti nel prevenire tanto la frammentazione in molti ipertesti sconnessi e proprietari quanto il pericolo che il Web finisca sotto il controllo monopolistico di una data società o di un solo stato;
2. L'adozione di standard per la codifica dei caratteri che permettono la presenza sul Web alle diverse lingue del mondo;
3. La creazione e la diffusione di standard per l'accessibilità, che tutelano chi è svantaggiato in termini di capacità (di vedere, sentire, muoversi e capire) e chi dispone di macchine con tecnologie obsolete.

La riflessione del padre dell'ipertesto globale si sposta dunque su un *topos* della teoria informatica, la dicotomia tra cervello e macchina e tra "rima" e "ragione". Se il cervello umano funziona tramite un complesso sistema di associazioni, e possiede la capacità di effettuare collegamenti che rientrino in un sistema coerente e consistente di conoscenze, le macchine, viceversa, hanno enormi capacità di calcolo, e possono processare l'informazione in forma di tabelle, vale a dire secondo una struttura gerarchica.

Per disporre di strumenti informativi più ricchi, l'architettura del Web deve pertanto definire con precisione i confini tra i "documenti" (i contenuti accessibili agli uomini) e i "dati"; una reale separazione tra documenti e dati permetterà alle macchine di processare le informazioni, una possibilità resa oggi complicata dal fatto che la gran parte dell'informa-

[49] Il W3C (acronimo di World Wide Web Consortium) è un organismo misto pubblico-privato con sede al MIT e nato allo scopo di promuovere e sviluppare tecnologie che garantiscano l'interoperabilità sul Web. Sul sito del W3C (<http://www.-w3.org/>) si trovano le specifiche e le linee guida del consorzio, oltre a software e a particolari applicazioni.

zione collegata in forma di ipertesto nella ragnatela condivisa "World Wide", contiene entrambi gli elementi mescolati[50].

Berners-Lee ha cura di precisare che non intende sostenere la possibilità di un'intelligenza artificiale - ne è conferma il presupposto su cui si fonda il suo ragionamento, la netta differenza tra il funzionamento del cervello umano e della macchina. Egli vuole semplicemente sottolineare i notevoli vantaggi che la possibilità di trattare informazione semanticamente strutturata reca con sé, favorendo l'interoperabilità, la portabilità e la durabilità dell'informazione (il riuso della conoscenza), e aprendo nuove possibilità di selezione del sapere.

Il discorso di Berners-Lee affronta un'altra questione centrale per la comunicazione scientifica: quella della selezione del sapere secondo criteri di "qualità". Se «è noto che una raccolta di testi, come un insieme di report tecnici o una biblioteca, include soltanto quegli articoli che raggiungano un certo livello di qualità» e «alcuni ritengono l'assenza di simili sistemi un limite del Web», tuttavia, precisa, è importante che il Web in sé non tenti di promuovere una singola nozione di qualità, ma continui a raccogliere ogni informazione, vera o falsa che sia. È questa una caratteristica che può certamente rappresentare un limite. Tuttavia, un'autorità centrale che esercitasse un controllo sulla qualità sarebbe assai più dannosa degli svantaggi che l'apertura del sistema può comportare; e se nessuno può essere in alcun modo obbligato a leggere letteratura di bassa qualità, è altresì vero che, come mostra la storia della scienza, appunti che oggi sono considerati marginali potrebbero, un domani, essere a fondamento di nuove idee dalla portata rivoluzionaria. Pertanto, il problema della selezione dell'informazione secondo criteri di qualità dev'essere così riformulato: «Come possiamo fornire all'utente la percezione soggettiva di qualità elevata, e allo stesso tempo mantenere il Web aperto a persone i cui criteri di giudizio sono diversi?». La questione, risponde il suo inventore, dovrà essere affrontata dotando i lettori di strumenti di filtro in grado di sfruttare l'organizzazione semantica dell'informazione, che trasformeranno i ricercatori in veri e propri "battitori di piste" i quali, come nell'esempio di Bush, potranno scambiarsi e condividere i percorsi e crearne di nuovi, mantenendo comunque l'ipertesto globale svincolato da una autorità centrale che decida quale informazione è appropriata, e per chi.

[50] La formattazione dei dati in HTML non permette che l'informazione sia pienamente interpretabile dalle macchine; è questo un limite strutturale del linguaggio, che non consente di separare pienamente il contenuto dalla sua struttura. Tuttavia, la nascita di nuovi formati di codifica (primo tra tutti XML, acronimo di *Extensible Markup Language*) e di nuove tecnologie di cui il W3C si è fatto promotore (che rientrano nel cosiddetto Web semantico) stanno favorendo l'accentuarsi di questa separazione.

Quest'ultimo punto, assieme agli altri discussi nel presente paragrafo, riguarda una questione più generale che è molto importante per la discussione qui proposta: quella dell'organizzazione della conoscenza o, in termini più astratti, dell'architettura dell'informazione. Su questo tema, un contributo può venire dai principi a fondamento dell'organizzazione della conoscenza emersi da secoli di ricerche, in primo luogo biblioteconomiche, cui l'idea e lo sviluppo del Web, come osservato nel secondo paragrafo, deve molto.

Se non è questa la sede per ricostruire la genesi della biblioteconomia, è utile tornare su alcuni passaggi della storia delle classificazioni per comprendere come il problema dell'organizzazione della conoscenza può essere affrontato nell'ambiente digitale e sullo spazio del Web. È cosa nota che l'origine della scienza moderna della classificazione si debba ai botanici, Konrad Gesner per primo e poi Carlo Linneo, che a metà del diciottesimo secolo creò un sistema di classificazione degli organismi viventi tutt'oggi utilizzato, basato su un grande albero gerarchico[51]. Le grandi classificazioni biblioteconomiche (la classificazione decimale di Dewey [CCD], la classificazione Decimale Universale [CDU] di Otlet e di La Fontaine, la Library of Congress Classification [LCC]) sono loro debitrici. Ma le svolte più significative per la ricerca e la selezione dell'informazione nel Novecento sono dovute a due studiosi: il primo è Charles Ammie Cutter, creatore dei Library of Congress Subject Headings, che ha affrontato la difficile questione delle ricerche per soggetto, stabilendo cinque regole convenzionali che rispecchiano l'impostazione filosofica dello stesso Berners-Lee: 1. uniformità: il concetto dev'essere formulato sempre nello stesso modo; 2. univocità: a un nome, deve corrispondere sempre lo stesso oggetto; 3. esaustività: il tema di un documento dev'essere descritto in modo completo; 4. specificità: un concetto dev'essere rappresentato in modo esatto dal linguaggio di indicizzazione; 5. predittività: il catalogo deve rispondere alle aspettative degli utenti[52].

Il secondo è Shiyali Ramamrita Ranganathan, matematico e bibliotecario indiano che negli anni 30 ha inventato il metodo dell'analisi a faccette.

> Una *faccetta* (in inglese "facet") è un particolare aspetto sotto il quale un argomento viene trattato; secondo Ranganathan, le faccette di qualsiasi classe si possono ricondurre a cinque categorie fondamentali: *personalità* (l'oggetto centrale di un discorso), *materia* (i componenti e le proprietà dell'oggetto), *energia* (le caratteristiche dinamiche dei processi che lo interessano), *spazio* (i suoi elementi geografici o in genere spaziali), e *tempo* (le sue fasi cronologiche). Con una classifi-

[51] P. Denis, G. Gohan, *Storia della biologia*, Einaudi, Torino 1999 (ed. orig. 1997).

[52] C. Gnoli, V. Marino, L. Rosati, *Organizzare la conoscenza. Dalle biblioteche all'architettura dell'informazione*, Tecniche Nuove, Milano 2006, p. 9.

cazione a faccette, il contenuto di un documento può essere descritto analiticamente nei suoi diversi aspetti; questi sono poi espressi tutti insieme, secondo una sequenza determinata da regole di funzionalità: perciò Ranganathan definisce questo tipo di classificazione *analitico-sintetica*[53].

Si tratta di un sistema rivoluzionario, anche se in pratica non molto applicato, perché permette di far rientrare un elemento in più categorie a differenza di quanto avviene nelle classificazioni ad albero. Si osservi anche, a margine, che Ranganathan non riteneva che i sistemi informatici avrebbero costituito uno strumento in grado di risolvere il problema della selezione dell'informazione *tout-court*. L'esempio che riportiamo sotto illustra il suo pensiero in proposito.

> Una questione importante è: possono i classificazionisti e i classificatori abdicare alla loro funzione e dipendere dal calcolatore, lasciando che esso prenda il loro posto nelle procedure che permettono al servizio bibliotecario di soddisfare tutte le cinque leggi della biblioteconomia? No, non possono. Perché la classificazione richiede il discernimento, discernimento del soggetto del documento, in tutte le faccette e le schiere che in esso si manifestano. Questo non si può fare con l'analisi statistica delle parole presenti nel documento, che è tutto quello che può fare la macchina.
> Un calcolatore può informare un lettore, quasi istantaneamente, della collocazione del libro di cui ha bisogno in quel momento, egli ne cita dettagli come l'autore e il titolo. Perché questo non comporta alcun discernimento. Ma l'esperienza insegna che la maggior parte dei lettori non sempre vuole o può cercare il suo documento per autore o per titolo. L'approccio più popolare è quello per soggetto in termini precisi, soprattutto se è un lettore impegnato in settori di frontiera della conoscenza [...]. Un paragone familiare è l'incapacità di una signora che desidera un nuovo sari (l'abito di seta colorata indossato dalle signore indiane) di specificarne con precisione il colore, il modello e gli altri dettagli. È raro che suo marito riesca a soddisfarla se è guidato unicamente dalle sue indicazioni verbali. Tuttavia, nel momento in cui il negoziante le dispiega davanti un assortimento di sari colorati, la signora è in grado di fare la sua scelta con piena soddisfazione senza alcuna perdita di tempo. Lo stesso succede con un lettore: il soggetto specifico di suo interesse spesso è ineffabile. Per questo, il browsing in una successione classificata di libri, o di voci di un catalogo, è una necessità.[54]

[53] C. Gnoli, *Il tavolino di Ranganathan*, «Bibliotime», III, 3, 2000, <http://www2.spbo.unibo.it/bibliotime/num-iii-3/gnoli.htm#faccette>.

[54] S.R. Ranganathan, *Prolegomena to Library Classification*. Asia Publishing House, New York, 1967 (ed. orig. 1937), <http://dlist.sir.arizona.edu/1151/>, XA 6, 61, <http://dlist.sir.arizona.edu/1151/23/ProlegomenaX.pdf>, tr. it. in C. Gnoli, V. Mari-

Tale affermazione può essere esaminata e discussa alla luce delle tecnologie oggi a disposizione, come cercheremo di fare nei prossimi capitoli. Qui tuttavia non stiamo considerando il Web come strumento, ma come spazio di documentazione. A tale proposito, può essere utile esplorare un'altra analogia, quella tra il libro (e, in senso generale, la biblioteca) e il nuovo spazio di documentazione inaugurato dal Web attraverso le cinque leggi della biblioteconomia che il bibliotecario indiano formulò nel 1931[55] e che sono considerate la base scientifica da cui poter dedurre le pratiche di funzionamento di ogni buona biblioteca:

1. *I libri sono fatti per essere usati.* I bibliotecari perciò devono promuovere l'accesso alla conoscenza e la disseminazione dei suoi prodotti.
2. *Ad ogni lettore il suo libro.* Pertanto, chiunque acceda a una biblioteca, indipendentemente dall'età, dal sesso, dalle capacità e dalla propensione a leggere, deve poter usufruire di tutti i libri in essa contenuti.
3. *Ad ogni libro il suo lettore*. È dunque compito dei bibliotecari fare in modo che un libro sia facilmente raggiungibile dal lettore potenzialmente interessato.
4. *Non far perdere tempo al lettore*. È essenziale che il servizio sia efficiente, sia per i lettori che per la biblioteca.
5. La biblioteca è un organismo che cresce. Si stabilisce così un'analogia tra la biblioteca e l'individuo, e anche una relazione con le evoluzioni tecnologiche che possono determinare una trasformazione della biblioteca stessa.

In un articolo recente, Alireza Noruzi ha proposto un'applicazione delle cinque leggi di Ranganathan al Web[56]:

1. *Le risorse sul Web sono fatte per essere usate*.

Questa legge implica il fatto che il Web debba poter essere usato e servire come strumento per l'apprendimento, e che l'informazione debba essere a disposizione di tutti: degli utenti, di chi fa ricerca e dell'intera società. Il fatto che sia stato pensato perché le risorse siano pubblicabili senza alcun filtro centralizzato, rispetta questa prima legge. Anche l'esistenza di strumenti come l'Internet Archive[57], un sito che indicizza tutte le pagine del Web nel corso del tempo, risponde a questa esigenza. Se una pagina scompare, è possibile rintracciarla tramite la *Wayback machine*, un motore di ricerca che attinge alle risorse non più attive.

2. *Ad ogni utente, la sua risorsa*.

La seconda legge ha diverse implicazioni importanti per il Web, rivelando «il bisogno fondamentale di un equilibrio tra la creazione di risor-

no, L. Rosati, *Organizzare la conoscenza*, cit., pp. 30-31.

[55] Cfr. S.R. Ranganathan, *Five Laws of Library Science*, Madras Library Association, Madras, Edward Goldston, London 1931, <http://dlist.sir.arizona.edu/1220/>.

[56] A. Noruzi, *Application of Ranganathan's Laws to the Web*, «Webology», 1, 2, 2004, <http://www.webology.ir/2004/v1n2/a8.html>.

[57] <http://www.archive.org/>.

se Web e il diritto basilare di tutti gli utenti ad avere accesso alle risorse Web di cui hanno bisogno, in ogni parte del mondo»[58]. È un aspetto che mette al centro la questione della diffusione e della disseminazione, sia in termini di politiche di accesso all'informazione (da parte degli esseri umani e dei motori di ricerca) resistenti alla censura, sia sul piano della costruzione di collezioni valide, esaustive e coerenti e pensate per tutti gli utenti e indipendentemente dal sesso, dall'età, dalla religione etc.

3. *Ad ogni risorsa, il suo utente*.

La terza legge afferma che una risorsa sul web esiste per ogni utente, e può essere tradotta in pratica tenendo a mente la necessità di connettere le risorse agli utenti, attraverso gli strumenti che il Web e la rete mettono a disposizione: *mailing list* e gruppi di discussione, indicizzazione nei motori di ricerca, *linking* a pagine già esistenti e pertinenti, costruzione di mappe strutturate dei siti web, che facilitino la navigazione e il reperimento delle informazioni.

4. *Non far perdere tempo all'utente*.

Questa legge è direttamente collegata alla precedente e costituisce una sfida per ogni progettista e sviluppatore. È fondamentale che il design di un sito sia pensato in modo da aiutare l'utente a trovare rapidamente l'informazione che cerca. «Questa quarta legge enfatizza un servizio efficiente per gli utenti, che implica una buona progettazione e una mappa/indice di un sito facile da capire»[59].

5. *Il Web è un organismo che cresce*.

L'ultima legge è abbastanza banale da comprendere. Abbiamo già osservato che pubblicare una pagina non necessita di alcun permesso. Perciò, quest'ultima legge costituisce una caratteristica vitale del Web e

> sottolinea la necessità di costanti aggiustamenti della nostra prospettiva nel rapportarci ad esso. Cambiamento e crescita vanno di pari passo, e richiedono flessibilità nella gestione delle collezioni, nell'uso del cyberspazio, nel mantenere e accrescere gli utenti e la natura dei programmi web. Le collezioni sul web aumentano e cambiano, così come cambiano le tecnologie dell'informazione e le stesse persone. Perciò, questa quinta legge riconosce che ci sarà indubbiamente una crescita e che dev'essere progettata in modo sistematico[60].

Si osservi in ultimo che sul Web l'utente sostituisce quello che nella biblioteca è il lettore. Utente è un termine che include il lettore, ma che comprende anche altri attori, primi tra tutti gli stessi computer. Per capire in che modo ciò possa e debba avvenire, è necessario considerare più da vicino i recenti sviluppi del Web.

[58] A. Noruzi, *Application of Ranganathan's Laws to the Web*, cit.

[59] *Ibidem*.

[60] *Ibidem*.

CAPITOLO 3

WEB DEI DATI E SOCIAL SOFTWARE

1. Cenni di topologia delle reti

Prima di introdurre ai recenti sviluppi del Web, è utile dare uno sguardo alla sua struttura da un punto di vista matematico. La conoscenza della topologia delle reti risulta infatti essenziale sia alla comprensione delle reti telematiche, sia all'uso dei metodi e degli strumenti, in primo luogo i motori di ricerca, che caratterizzano la professione del ricercatore nell'epoca di Internet.

Da un punto di vista topologico, sia Internet che il web sono grafi, vale a dire insiemi di vertici uniti da archi. La teoria dei grafi è stata inaugurata da Eulero nel 1736[1] per risolvere un problema pratico che affliggeva gli abitanti di Königsberg. La città di Königsberg ha una geografia particolare: si trova alla confluenza di due fiumi, comprende un isolotto ed è divisa in quattro parti che al tempo di Eulero erano unite tramite sette ponti. Si narra che gli abitanti si divertissero a scommettere sulla possibilità di trovare un percorso che, partendo da una qualsiasi delle quattro zone della città, permettesse loro di attraversare ciascun ponte soltanto una volta, ritornando infine al punto di partenza. Il matematico rappresentò ciascuna delle quattro zone della città con un cerchio ("vertice"), indicando ogni ponte con una linea ("arco") e fornì una dimostrazione matematica dell'impossibilità di trovare un tale percorso[2] dando così origine alla cosiddetta teoria dei grafi (cfr. la figura 1 nella pagina seguente).

Il problema fu concretamente risolto, e, per buona pace dei suoi abitanti, nel 1875 a Königsberg fu costruito un ottavo ponte.

La teoria inaugurata da Eulero è oggi considerata il fondamento della attuale concezione delle reti. Vale a dire che «nella loro architettura, i

[1] Eulero, *Solutio problematis ad geometriam situs pertinentis*, 1736, <http://math.-dartmouth.edu/~euler/docs/originals/E053.pdf>.

[2] Eulero basò la sua dimostrazione sul fatto che i vertici con un numero dispari di spigoli debbono trovarsi al principio o al termine del percorso, e che un percorso che comincia in un punto e finisce in un altro non può avere più di due nodi siffatti. È facile osservare che il grafo di Königsberg ha quattro nodi con un numero dispari di collegamenti e che quindi tale percorso non esiste.

Francesca Di Donato, *La scienza e la rete : l'uso pubblico della ragione nell'età del Web*,
ISBN 978-88-8453-494-1 (print) ISBN 978-88-8453-500-9 (online)

grafi o le reti nascondono proprietà che possono limitare o favorire ciò che possiamo fare con loro»[3].

Figura 1. Il disegno originale di Eulero

In un saggio recente, il fisico rumeno Albert-László Barabási ripercorre le tappe essenziali nella storia della scienza delle reti, e definisce un particolare modello teorico, le reti a invarianza di scala, di cui mostra esempi significativi. Internet, il World Wide Web, la rete delle citazioni scientifiche, le presenze sul set degli attori di Hollywood e persino la diffusione dei virus hanno una topologia comune, ovvero sono riconducibili a quest'unico modello. Sono reti che obbediscono a leggi comuni.

Si tratta di reti distribuite, dinamiche e in crescita, tenute insieme da una gerarchia di connettori, che formano una tela senza il ragno, vale a dire autorganizzata. Che cosa significa, esattamente? È possibile dare una risposta a questa domanda analizzando alcune caratteristiche di queste reti anche senza entrare in complesse definizioni matematiche.

1. Il Web è un grafo (i cui vertici si chiamano “nodi” e gli archi “link”) orientato, cioè una rete diretta i cui link sono monodirezionali, vale a dire che sono orientati in un'unica direzione. In pratica, ciò significa che nel momento in cui inserisco un link ipertestuale in una pagina Web, io stabilisco un collegamento dalla mia pagina alla pagina che cito, in modo analogo a quanto avviene in un libro, in cui le citazioni vanno in un'unica direzione, verso scritti pubblicati in precedenza.

2. Il Web è un “mondo piccolo”, in cui il numero di passaggi necessari a raggiungere un punto qualunque del suo spazio è molto basso. La teoria dei mondi piccoli, nota anche come principio dei “sei gradi di separazione”, è stata definita dal sociologo americano Stanley Milgram come esito di un esperimento condotto nel 1961. Milgram si propose di ri-

[3] A.L. Barabási, *Link. La scienza delle reti*, Einaudi, Torino 2004, p. 14.

spondere alla questione: qual è la distanza tra due cittadini qualsiasi degli Stati Uniti? In altri termini: quanti collegamenti (link) sono necessari per connettere tra loro due individui che non si conoscono scelti a caso? Selezionò due destinatari finali (la moglie di uno studente di teologia nel Massachussetts e un agente di cambio di Boston) e scelse casualmente alcuni abitanti di piccole città del Kansas e del Nebraska, a cui inviò una lettera in cui spiegò gli obiettivi dell'esperimento e le istruzioni da seguire[4]. Il risultato fu sorprendente: in media, negli Stati Uniti, il numero minimo di intermediari necessari per collegare due sconosciuti qualsiasi attraverso persone di loro conoscenza è circa sei. Vale a dire che, più in generale, la società è una rete di sei miliardi di nodi in cui la distanza media tra un nodo e l'altro non supera una manciata di link.

È stato dimostrato che tale proprietà è scientificamente fondata e valida in diverse reti a invarianza di scala. Nel 2004, i gradi di separazione del Web erano diciannove. Questa caratteristica è propria anche della rete delle citazioni scientifiche, come mostra il numero di Erdos[5]. Paul Erdos era un matematico noto per aver collaborato con moltissimi scienziati e la cui produzione scientifica è stata molto vasta. Il numero a lui intitolato misura la distanza dei membri della comunità scientifica da Erdos, assegnando il numero 1 ai suoi coautori, cioè a coloro che hanno firmato con lui un testo, il 2 ai coautori dei suoi coautori, e via dicendo. I gradi di separazione tra Erdos e gli scienziati appartenenti al suo ambito disciplinare sono all'incirca sei. Si osservi però che Erdos (come Kevin

[4] Milgram inviò la lettera seguente:

«COME PRENDERE PARTE A QUESTO STUDIO

1) AGGIUNGETE IL VOSTRO NOME ALLA LISTA CHE TROVATE IN FONDO A QUESTO FOGLIO, affinché chi riceve per primo la lettera possa sapere da chi proviene.

2) STACCATE UNA CARTOLINA POSTALE, COMPILATELA E RISPEDITELA ALL'UNIVERSITA' DI HARVARD. L'affrancatura non è necessaria. La cartolina è molto importante: ci permetterà di seguire le tracce del documento nel suo viaggio verso il destinatario finale.

3) SE CONOSCETE DI PERSONA IL DESTINATARIO FINALE, SPEDITEGLI/LE DIRETTAMENTE IL DOCUMENTO. Fatelo soltanto se lo avete già incontrato in precedenza, e se vi date del tu.

4) SE NON CONOSCETE DI PERSONA IL DESTINATARIO FINALE, NON CERCATE DI CONTATTARLO DIRETTAMENTE. SPEDITE INVECE QUESTO DOCUMENTO (COMPLETO DI CARTOLINA POSTALE) A UN VOSTRO CONOSCENTE CHE RITENETE ABBIA MAGGIORI PROBABILITA' DI CONOSCERE IL DESTINATARIO FINALE. Potete spedirlo a un amico, a un parente o a un conoscente, ma dev'essere qualcuno a cui date del tu». Cit. in A.L. Barabási, *Link. La scienza delle reti*, cit., pp. 31-32. Si veda S. Milgram, *The "Small World" Problem*, «Psychology Today», 1, 1967, pp. 60-67.

[5] Sulla storia e le origini del numero di Erdos, si veda la relativa voce di wikipedia <http://it.wikipedia.org/wiki/Paul_Erd%C5%91s> (10/09). Cfr. anche C. Goffman, *And what is your Erdos number?*, «American Mathematical Monthly», 76, 1969.

Bacon nel mondo degli attori del cinema) ha alcune caratteristiche che lo rendono un nodo "particolare" nella rete.

3. Il web è caratterizzato dalla presenza di connettori *(hub)* e di autorità (*authorithy*). I primi hanno un alto numero di link in uscita, puntano cioè a un numero elevato di risorse, come ad esempio i siti di ebay o Amazon; le seconde, viceversa, contengono molti link in entrata, cioè sono altamente citate (ne sono esempi i siti di istituzioni importanti o le homepage dei quotidiani principali). L'alto numero di link è misura della loro popolarità. Molti connettori sono anche autorità, per esempio Google. L'architettura del Web è dominata da pochissimi nodi altamente connessi (detti appunto *hub)* che tengono insieme molti nodi poco connessi.
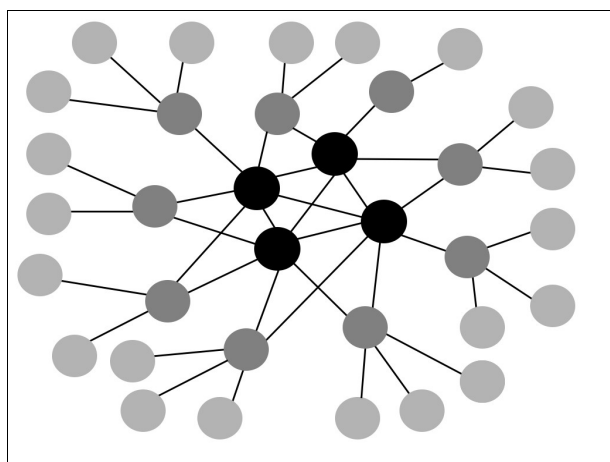
Figura 2. Pochi nodi altamente connessi (al centro) sono collegati alla maggior parte degli altri nodi, debolmente connessi.

Tale caratteristica accomuna il Web anche alla rete della società umana: «Disseminate in varie occupazioni, esiste una manciata di persone che possiede l'abilità davvero straordinaria di stringere un numero eccezionale di amicizie e conoscenze. Sono i connettori»[6]. I connettori sono presenti in tutte le grandi reti complesse e svolgono un ruolo determinante. Sul Web, essi fungono da vetrine privilegiate. Rimanendo nell'ambito della comunicazione scientifica, se consideriamo le risorse sulla rete al pari di libri e riviste, si può dire che i connettori agiscono come biblioteche prestigiose o, meglio ancora, grandi librerie. Se una risorsa compa-

[6] M. Gladwell, *Il punto critico. I grandi effetti dei piccoli cambiamenti*, Rizzoli, Milano 2000, cit. in A.L. Barabási, *Link. La scienza delle reti*, cit., p. 61.

re nella prima pagina in una *query* su Google è, per restare ai libri, come finire nella vetrina di un enorme negozio di una grande città: rende il nostro documento estremamente visibile[7].

4. Una caratteristica, piuttosto intuitiva, che deriva dalla precedente, è che sul Web il numero di link dà la misura di visibilità di un nodo. Maggiori sono i link da e verso una pagina, più la pagina è visibile.

5. Il Web è caratterizzato dalla legge 80/20, nota anche come principio di Pareto, che è sintetizzabile nell'affermazione secondo cui, su grandi numeri, la maggior parte degli effetti (indicativamente l'80%) è dovuta a un numero ristretto di cause (indicativamente, il 20%). Pareto dimostrò infatti che in una determinata regione solo pochi individui possedevano la maggior parte della ricchezza. La legge è stata tuttavia verificata grazie a osservazioni empiriche di numerosi fenomeni. Essa è dunque applicabile sia al Web, in cui l'80% di link proviene dal 20% delle pagine, sia alla rete della comunicazione scientifica tradizionale, in cui, come osservava De Solla Price, l'80% citazioni proviene dal 20% degli scritti.

6. Il Web è una rete a invarianza di scala, i cui nodi sono distribuiti secondo leggi di potenza in base a due princìpi: il principio della crescita (sempre nuovi nodi vengono aggiunti progressivamente nel tempo) e il principio del collegamento preferenziale (è un comportamento comune collegare le proprie pagine a pagine che hanno molti link, che sono cioè altamente connesse).

A causa di tali caratteristiche e proprietà, dunque, il Web non forma una singola rete omogenea ed è costituito da numerosi "sottomondi", in modo analogo a quanto avviene per la rete delle citazioni scientifiche. Il suo spazio è diviso in quattro continenti di dimensioni pressoché equivalenti come rappresentato in figura 3.

Il primo è il "corpo centrale", comprende circa il 25% delle pagine, ed è facile da navigare perché molto interconnesso. Esso contiene i grandi connettori come i motori di ricerca (primo tra tutti google) e i grandi siti altamente linkati e visitati (per fare alcuni esempi: Amazon.com, ebay, youtube) ed è la zona più visibile del Web. Il secondo, detto "continente *in*", è più difficile da navigare poiché consente di muoversi verso il "corpo centrale" ma non viceversa (come accade al pesce in una nassa). Allo stesso modo, il terzo continente (detto "continente *out*") è raggiungibile dai nodi del corpo centrale, ma una volta usciti da tale zona non ci sono link per tornare indietro (esso contiene prevalentemente siti aziendali o gli archivi dei quotidiani online). Il quarto continente è fatto di tentacoli e isole separati, gruppi di pagine collegate tra loro ma non al "corpo centrale". Si osservi inoltre, per inciso, che il Web gode delle proprietà dei

[7] Usa questo paragone Vincenzo Letta nel suo *In Google we trust. Storia, vizi e virtù di un motore di ricerca nordamericano*, Tesi di laurea in etica della comunicazione, Facoltà di Lettere e filosofia dell'Università di Pisa, 2008, <http://etd.adm.unipi.it//theses/available/etd-10132008-104502/>, p. 5.

frattali, vale a dire che, visto da lontano o da vicino, ha la medesima struttura[8].
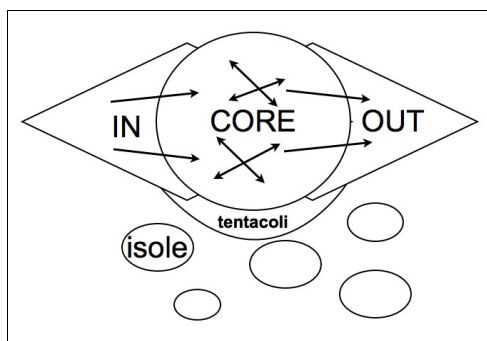
Figura 3. I quattro continenti del Web

Recenti studi hanno stimato che soltanto metà del Web è navigabile, cioè il "corpo centrale" e il "continente *out*". Questa caratteristica è una sua proprietà topologica: vale a dire che la possibilità di trovare un percorso tra due nodi è indipendente dalle capacità dell'uomo e dei migliori motori di ricerca, esistenti o possibili (proprio come a Königsberg).

Non solo: la struttura che abbiamo descritto riguarda la parte visibile del Web, ma il Web è molto più vasto e più ricco. Esiste infatti un Web sommerso *(deep Web)* che è introvabile dai motori e dai *crawler*[9] (e che ad esempio contiene le pagine dinamiche create in risposta a una *query* su un database, o pagine accessibili tramite password), le cui dimensioni sono stimate essere tra le 400 e le 550 volte quelle del Web visibile[10].

[8] Il frattale è un oggetto geometrico che si ripete nella sua struttura allo stesso modo su scale diverse (esempi di frattali in natura sono i fiocchi di neve, gli alberi, i cavolfiori). Per una definizione più completa ma divulgativa, si veda la voce di wikipedia all'URL <http://it.wikipedia.org/wiki/Frattale> (10/09).

[9]Un *crawler* (detto anche *spider* o robot), è un software che analizza i contenuti di una rete (o di un database) in un modo metodico e automatizzato, in genere per conto di un motore di ricerca. Cfr. <http://it.wikipedia.org/wiki/Crawler> (10/09).

[10] Cfr. I.H. Witten, M. Gori, T. Numerico, *Web Dragons. Inside the Myths of Search Engine Technology*, Morgan Kaufmann Publishers-Elsevier, San Francisco 2007, pp. 96-7. Il testo è un utilissimo strumento per la ricerca sul Web (assieme al sito Searchlores di Fravia <http://www.searchlores.org/>), ed è in buona parte complementare a questo libro. L'impostazione filosofica alla base di *Web Dragons* ci sembra tuttavia differente dalla nostra, in quanto orientata da un approccio che si basa sulle teorie dell'intelligenza artificiale, da cui i sostenitori del Web semantico prendono invece le distanze. Un punto, questo, che sarà meglio approfondito nei futuri sviluppi di questa ricerca.

Ne fanno parte, per restare alla letteratura scientifica, anche le biblioteche digitali e la cosiddetta “letteratura grigia”, in cui rientrano tesi di laurea e di dottorato, rapporti, relazioni e presentazioni a convegni, saggi in attesa di accettazione da parte di riviste accademiche, cataloghi, manuali di prodotti hardware e software, dispense universitarie[11].

Si osservi dunque che essere online non significa, di per sé, essere visibili. Il modo preferenziale in cui si naviga sul Web è attraverso i motori di ricerca, che però pescano i risultati solo all'interno del corpo centrale e nel continente out del Web “di superficie”.

In che modo dunque i motori trovano le pagine sul Web?

Google, il motore di ricerca più utilizzato nel mondo, è stato inventato da due giovani ricercatori di teoria dei grafi, Sergej Brin e Larry Page, mentre analizzavano le caratteristiche dei grafi orientati (i cui collegamenti, ricordiamo, sono unidirezionali). Brin e Page concentrano l'analisi su un problema particolare: come tornare indietro in un grafo orientato, cioè come scoprire se un sito è stato linkato, e da chi. Nel secondo capitolo si è detto che Berners-Lee, perso il suo primo programma *Enquire* che permetteva di costruire reti indirette con link bidirezionali, lo ricostruisce con link che vanno in una sola direzione. Brin e Page ne raccolgono la sfida e danno vita a *BackRub,* letteralmente “massaggio alla schiena”, che diventa poi *PageRank*, l'algoritmo che consente a Google di trovare le risorse sfruttando questo limite architettonico del Web e trasformandolo in una ricchezza[12]. All'origine della loro invenzione sta l'intenzione di scandagliare il Web, impresa in cui, come ricorda Barabási, non sono gli unici a cimentarsi. Page decide di farlo cercando di capire dove è possibile arrivare a partire da una pagina, e poi invertendo la direzione.

> Adesso aveva chiara una nuova teoria: conteggiare il numero di link che puntava a un sito Web voleva dire misurare la popolarità di quel sito e si accorse subito che, da questo punto di vista, il link assomigliava a un altro concetto che viveva dentro il mondo accademico, la citazione. Essa costituiva il criterio base utilizzato nel sistema di valutazione delle pubblicazioni universitarie. Al mero criterio quantitativo ebbe modo così di capire che si poteva aggiungerne anche uno di autorevolezza. Ma se anche al link di provenienza doveva essere applicato lo stesso criterio per valutarne l'importanza, la cosa costituiva una difficile sfida di matematica ricorsiva. I calcoli diventavano complicati e si ingigantivano a dismisura[13].

[11] Cfr. F. Metitieri, R. Ridi, *Biblioteche in rete. Istruzioni per l'uso*, Laterza, Roma-Bari 2005, in part. Capitolo 10, <http://www.laterza.it/bibliotecheinrete/Cap10/Cap10_10.htm>.

[12] S. Brin, L. Page, *The Anatomy of a Large-Scale Hypertextual. Web Search Engine*, 2000, <http://infolab.stanford.edu/~backrub/google.html>.

Il giovane dottorando americano affronta lo stesso problema fronteggiato da Eugene Garfield pochi decenni prima, ma con una difficoltà in più, che lo rende a prima vista irrisolvibile. Lo spazio del Web, a differenza del database dello *Science Citation Index*, è aperto e in continua crescita. Così, assieme a Brin, Page crea un sistema per valutare la popolarità di un sito, incrociando criteri quantitativi e criteri di autorevolezza, cioè pesando i link sulla base della loro importanza. In breve, Google identifica la posizione di un nodo nel grafo e privilegia, nelle risposte alle *query*, le grandi autorità e i grandi connettori. Ad esempio, è intuitivo che se la mia pagina personale viene linkata da quella di un sito di un mio collega, secondo Google l'importanza di quel collegamento è minore rispetto a quella di un link dal sito di repubblica.it, anche se il ricercatore che mi cita è uno scienziato illustre. Dunque ci sono link che pesano più di altri in termini di popolarità. In una riflessione sulla comunicazione scientifica non deve tuttavia sfuggirci che questo tipo di peso non necessariamente misura il valore scientifico del collegamento.

Riprenderemo questo argomento nel prossimo capitolo. Basti qui sottolineare che PageRank, pur agendo sull'architettura del Web, ne altera alcune caratteristiche. In primo luogo, sfrutta la possibilità di percorrere i link a ritroso, trasformando il nucleo centrale di un grafo orientato in una rete indiretta. Ma c'è un punto ulteriore, forse ancora più rilevante. L'algoritmo di Google è proprietario e segreto – a differenza delle tecnologie grazie alle quali il Web è nato e si è sviluppato, che sono aperte e di pubblico dominio.

2. Oltre i limiti del Web: il Web semantico

Sappiamo dunque che è possibile navigare soltanto in una frazione del Web, vale a dire che la maggior parte delle sue pagine può essere rintracciata soltanto conoscendone l'indirizzo (URL o URI) e che solo una minima parte di esse è raggiungibile a partire da altri nodi della ragnatela. Ma come giungiamo a un sito che ci interessa? Abbiamo visto che lo strumento essenziale a reperire le informazioni contenute nelle pagine Web sono i motori di ricerca. Digitando una o più parole su Google, il sito di Google ci restituisce un elenco di link a pagine che contengono informazioni pertinenti ai fini della nostra ricerca. Tuttavia, non sempre le informazioni che troviamo corrispondono a quanto cerchiamo; viceversa, di solito la parte maggiore dei risultati è decisamente "fuori tema". Infatti, se i motori di ricerca si sono dimostrati utilissimi nel setacciare rapidamente un numero molto elevato di pagine, non si rivelano altrettanto utili nel valutare la qualità del documento. Assieme a informazioni

[13] V. Letta, *In Google we trust. Storia, vizi e virtù di un motore di ricerca nordamericano*, cit. p. 16.

rilevanti, i motori ci restituiscono moltissimi link a pagine il cui contenuto è estraneo a quanto cerchiamo. Un problema che non dipende dalle caratteristiche topologiche della rete, ma da un suo limite architettonico.

In pratica, le difficoltà che incontrano i motori di ricerca dipendono dal fatto che questi si limitano a controllare la presenza nei documenti di determinati termini (parole chiave), un indizio che dice molto poco sul reale contenuto delle pagine che troviamo[14]. Perché dunque non si investe nel migliorare le tecniche e gli algoritmi di ricerca? In parte, lo si fa. Tuttavia, l'architettura originaria del Web presenta alcuni limiti che non dipendono da un difetto negli algoritmi di ricerca utilizzati dai motori. Il limite principale consiste piuttosto nel fatto che, sul Web, le diverse informazioni di un documento HTML sono mescolate al suo interno e non sono strutturate semanticamente, rendendo impossibile alle macchine trattare i dati sparsi nelle pagine secondo il significato che attribuiamo loro nel contesto in cui sono inseriti - mentre il significato dei dati all'interno del contesto è essenziale in qualsiasi ricerca.

Prendiamo ad esempio il caso in cui io decida di vendere la mia automobile tramite un annuncio in rete. L'annuncio che decido di postare in una pagina Web è più o meno del tipo "Vendo Toyota Yaris nera, ottimo stato, buon prezzo, Pisa, Italia". Tuttavia, se cerco "Yaris nera in vendita a Pisa, Italia", i motori di ricerca mi restituiranno un elenco sterminato di pagine che contengono quelle parole, ma il cui contenuto, nella maggior parte dei casi, sarà molto lontano da quello che desidero, informazione che viene definita *noise*, "chiasso", poiché distoglie la mia attenzione dai risultati che effettivamente mi interessano. Otterrò invece risultati migliori decidendo di inserire la mia offerta in un sito specializzato che contiene un modulo i cui campi di ricerca siano predefiniti e indichino, ad esempio, la casa automobilistica, il modello, l'anno di immatricolazione dell'auto e altri dati rilevanti ai fini della ricerca. Questo perché la pagina è scritta in un formato leggibile da una macchina, la quale sarà in grado di identificare e conservare il significato delle singole parti del documento.

Nel processo di ricerca non conta soltanto trovare la risposta a una domanda; ad assumere un rilievo fondamentale sono le strutture in cui è inserita la singola informazione. Il progetto originario del Web prevedeva in effetti che tali strutture fossero in qualche modo visibili e rintracciabili tanto dagli umani quanto dai computer[15]. Viceversa, sul Web come

[14] T. Berners-Lee, *L'architettura del nuovo Web*, cit., p. 155.

[15] Si osservi con attenzione la figura 1 nella prima proposta presentata da Berners-Lee al CERN (*Information Management: A proposal*, marzo 1989-maggio 1990, <http://www.w3.org/History/1989/proposal.html>). Nella figura, gli archi che collegano le singole pagine (e che corrispondono ai link) sono contrassegnati da "etichette" che definiscono il tipo di collegamento che esiste tra due risorse. I tipi di collegamento possono essere diversi: ad esempio "Tim Berners-Lee" "ha scritto" "questo documento". "Questo documento" "descrive" "l'ipertesto". Le relazioni "ha scritto" e "de-

è stato implementato e come lo conosciamo, i link non sono "etichettati". Questa caratteristica ne indebolisce le potenzialità.

Dal 1999 il World Wide Web Consortium (W3C)[16], il consorzio fondato e diretto da Tim Berners-Lee allo scopo di promuovere standard che assicurino l'interoperabilità del Web, ha concentrato le ricerche sui modi in cui risolvere il problema, dando vita a un orientamento noto come Web semantico[17]. Scopo di tale orientamento è progettare e sviluppare una estensione del Web che lo trasformi in una ragnatela di dati elaborabili dalle macchine. La scommessa fatta propria dal W3C è far diventare la rete in grado di interpretare le nostre richieste. Il web diviene semantico nel momento in cui, nella rete di collegamenti tra dati di natura differente ed espressi in forma diversa, i computer diventano in grado di trattare i dati in modo da inferire nuova conoscenza a partire da quella nota. Da un punto di vista tecnico, ciò è possibile strutturando l'informazione in modo tale che i documenti non restino "isole di dati" ma diventino "data base aperti" da cui un programma possa attingere informazioni.

Per comprendere il senso di tale affermazione, è utile introdurre la differenza tra "*information retrieval*" (recupero di informazione) e "*data retrieval*" (risposta automatizzata alle domande). Obiettivo della prima è produrre documenti che sono rilevanti per una *query*; questi documenti non devono essere unici, e interrogazioni successive possono dare luogo a risultati completamente diversi. Obiettivo del secondo è invece produrre la risposta corretta a una domanda. Il Web semantico si occupa di quest'ultimo aspetto[18]. Ma che cosa significa Web semantico, in pratica? E come ci apparirà il nuovo Web? Una prima risposta si trova nell'esempio che segue:

> Quando il telefono squillò, il sistema di intrattenimento stava cantando a squarciagola "We Can Work It Out" dei Beatles. Nel momento in cui Pete alzò la cornetta, il suo telefono abbassò il volume mandando un messaggio a tutti i dispositivi *locali* con un *controllo del volume*. All'altro capo della linea c'era sua sorella, Lucy, dallo studio medico: "la mamma ha bisogno di una visita specialistica e poi dovrà eseguire una serie di trattamenti. Due volte a settimana, o giù di lì. Ora chiedo al mio agente di fissare gli appuntamenti". Pete accettò subito di condividere l'impegno.

scrive" si traducono banalmente in link. Ma sono link dal significato diverso; un significato che, nell'idea originaria di Web, veniva esplicitato.

16 Il sito del W3C è all'URL <http://www.w3.org>.

17 Si veda in particolare questa sezione del sito del W3C <http://www.w3.org/2001/sw/>.

18 T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, *A Framework for Web Science*, cit., p. 18.

> Dallo studio del medico, Lucy istruì il suo agente semantico attraverso il browser del suo palmare. L'agente di Lucy trovò immediatamente dall'agente del medico le informazioni relative al *trattamento prescritto*, controllò diverse liste di *offerte* e controllò quelle che *rientravano nel piano assicurativo* della madre *nel raggio di 20 miglia* da casa sua, e con un *rating* di fiducia *eccellente* o *molto buono*. Poi l'agente cominciò a provare a cercare gli *appuntamenti liberi* (offerti dagli agenti dei singoli provider tramite i loro siti Web) compatibili con le disponibilità di Lucy e di Pete (le parole in corsivo indicano i termini la cui semantica, o significato, vengono definiti per gli agenti attraverso il Web semantico).
> In pochi minuti, l'agente presentò loro un programma. Pete non lo apprezzò. Lo University Hospital si trovava dalla parte opposta della città rispetto alla casa della madre, e avrebbe dovuto tornare indietro nel caos del traffico dell'ora di punta. Programmò il suo agente in modo da rifare la ricerca con preferenza più strette su *luogo* e *orario*. L'agente di Lucy, avendo completa fiducia in quello di Pete riguardo a questa specifica ricerca, fornì a quello assistenza immediata fornendogli l'accesso ai certificati e agli *shortcut* dei dati che aveva trovato.
> Quasi istantaneamente fu presentato un nuovo programma: la clinica era molto più vicina e gli orari anticipati, ma c'erano due avvisi. In primo luogo, Pete avrebbe dovuto spostare due appuntamenti poco importanti. Controllò di che cosa si trattava, e vide che non era un problema. L'altro riguardava il fatto che la clinica non rientrava nell'elenco di *terapeuti* coperti dalla compagnia di assicurazione: l'agente lo rassicurò "Servizio coperto e piano assicurativo verificato sicuro da altri mezzi" "(Dettagli?)"
> Lucy si rese assente nello stesso istante in cui Pete brontolava: "risparmiami i dettagli" e tutto fu risolto[19].

Abbiamo osservato come, da un punto di vista tecnico, l'architettura originale del Web si basasse esclusivamente su tre princìpi semplici: URL (la possibilità di puntare a una risorsa dandole un nome dal significato univoco), HTTP (il protocollo di trasporto delle pagine sul Web) e HTML (il linguaggio di codifica delle pagine Web, che permette di inserire link). Per far sì che i computer possano essere in grado di eseguire compiti come quello descritto, si sono rese necessarie alcune trasformazioni nell'architettura del Web. Che cosa significa strutturare semanticamente l'informazione? E come si può farlo?

È possibile costruire una risposta a entrambe le domande considerando da vicino le trasformazioni tecnologiche sulle quali si fonda il Web semantico, oggi detto anche Web dei dati. Da un punto di vista architetto-

[19] T. Berners-Lee, J. Hendler, O. Lassila, *The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities,* «Scientific American Magazine», 2001, <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>.

nico, il Web semantico è stato rappresentato nel 2001 da Tim Berners-Lee[20] come una piramide di sette strati e composta da nove elementi, una piramide che "sorregge" tre tipologie di informazioni ("documenti auto-descrittivi", "dati" e "regole").
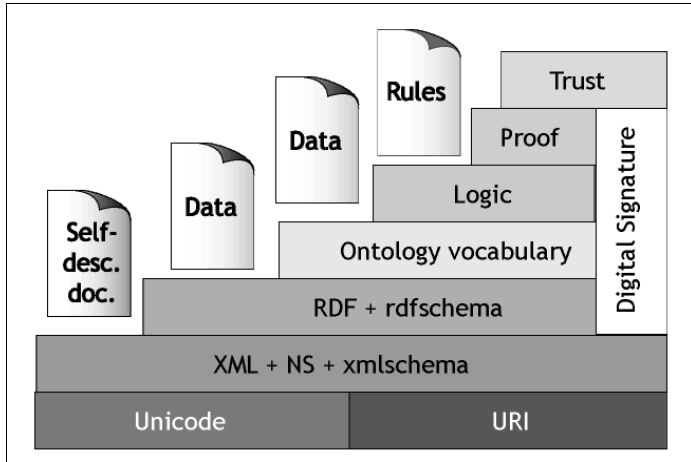
Figura 4. Semantic web Layers Cake

Il significato della nuova architettura dev'essere spiegato a partire dalla base della piramide. Gli elementi in basso sorreggono gli strati più alti, e ne sono una precondizione. Infatti, il rapporto tra un piano della piramide e quello successivo può essere interpretato come un rapporto di condizionamento inteso in senso kantiano, vale a dire che il web può essere concepito come un insieme di strati con standard, linguaggi o protocolli che agiscono come piattaforme sulle quali possono poggiarsi formalismi nuovi, più ricchi e più espressivi. Piattaforme che sono intese più neutrali possibile, in cui l'architettura a strati ha funzione di regolamentazione e non è prescrittiva[21]. Un'analogia che può aiutare a comprendere il tipo di rapporto che esiste tra i sette strati della "torta semantica" è quella con il funzionamento di un aeroporto, dove la partenza e l'arrivo

[20] La slide è disponibile online all'URL <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>.

[21] Gli studiosi di Web *science* sostengono che il rapporto tra gli strati si possa definire come *supervenienza,* termine che in filosofia indica un modo per spiegare la generazione di significati per cui un discorso A è superveniente a un discorso B se un cambiamento in A comporta un cambiamento in B, ma non viceversa. T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, *A Framework for Web Science*, cit., p. 72.

di ogni aereo dipendono dalla corretta concatenazione ed esecuzione di una serie di procedure standard (dal *check-in* dei passeggeri al loro sbarco, dalla spedizione alla riconsegna dei bagagli, dalle fasi di decollo e di atterraggio, seguite dai piloti e monitorate dalle torri di controllo, etc...).

È inoltre importante osservare che, nel passaggio dagli strati inferiori a quelli superiori (in particolare gli ultimi due, "*Proof*" e "*Trust*"), le innovazioni che sono introdotte sono poco definite sul piano tecnologico. Ciò che viene invece illustrato è piuttosto l'impatto socio-culturale che le trasformazioni in atto nell'architettura del Web saranno in grado di produrre nelle comunità che operano sul Web, un aspetto di particolare interesse ai fini del nostro discorso.

Un'analisi dettagliata dell'architettura del Web semantico esula dallo scopo di questo contributo. Qui ci limitiamo a riassumere gli aspetti essenziali delle innovazioni tecnologiche proposte, che possiamo concentrare in quattro punti:

a) *Metadati* ("dati sui dati"): servono a descrivere o annotare una risorsa in modo da renderla maggiormente comprensibile agli utenti. Normalmente i metadati sono di natura descrittiva e includono informazioni relative a un documento tra cui ad esempio l'autore, il titolo o l'*abstract*, il tipo di file, i diritti d'accesso o il numero di versione. L'aggiunta di dati sui dati è utile per organizzare le risorse, per l'archiviazione e per identificare l'informazione. I metadati sono dunque una nuova forma di catalogazione, che può essere inserita nei documenti digitali.

> Uno schema di metadati può essere creato da chiunque, a seconda della natura dei propri documenti e dell'uso che si prevede di farne. Esistono infatti schemi di metadati sviluppati per particolari settori, come i testi letterari (TEI), le informazioni geospaziali (NSDI), gli stessi cataloghi bibliografici (MARC)[22].

Tuttavia, la funzione più importante dei metadati è promuovere l'interoperabilità, consentendo la combinazione di risorse eterogenee tra piattaforme diverse senza perdere informazioni rilevanti[23]. Schemi di metadati per il Web sono ad esempio quelli della *Dublin Core Metadata Initiative*, un sistema costituito da un nucleo di 15 elementi essenziali ai

[22] C. Gnoli, V. Marino, L. Rosati, *Organizzare la conoscenza. Dalle biblioteche all'architettura dell'informazione*, cit., p. 60.

[23] Per l'uso di chi naviga sul Web, i metadati possono essere non-strutturati; viceversa, per essere comprensibili alle macchine è necessario che tali informazioni siano strutturate. RDF offre meccanismi per integrare schemi di metadati. Cfr. T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, A *Framework for Web Science*, cit., pp. 36-39.

fini della descrizione di qualsiasi materiale digitale accessibile tramite la rete[24].

Si parla dunque di "*open data*" per definire un'importante caratteristica del web semantico, cioè la disponibilità dei dati e la conseguente possibilità di identificarli e citarli. Il Web semantico è perciò un'estensione del Web tradizionale nel senso che è il successivo passaggio nel linking ed è pensato per funzionare nel contesto di un modello relazionale di dati, in cui il link, da collegamento generico e cieco tra due documenti, diviene capace di esprimere relazioni concettuali, che convogliano significati. In pratica, sul Web ogni "dato", ogni pezzetto di informazione, viene identificato da un URI. Perciò, si usa dire che sul web semantico tutto è un URI.

b) Gli URI (identificatori di dati) vengono messi in relazione tra loro tramite il linguaggio *Resource Description Framework* (RDF), che consente di costruire asserzioni tramite triple formate da soggetto predicato e oggetto e di collegare tra loro le triple in un unico grafo (in cui soggetto e oggetto sono nodi, URI, e il predicato è il link). Si vedano le figure che seguono, per un esempio. La prima (figura 5) mostra come si costruiscono le triple.
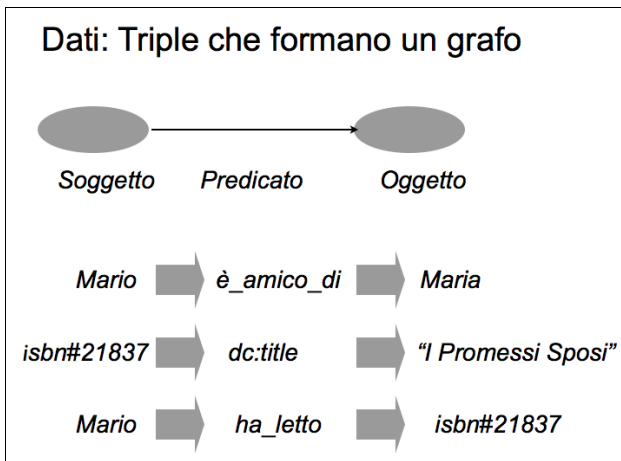
Figura 5. Costruzione di triple soggetto → predicato → oggetto

La seconda (figura 6) mostra come si costruiscono grafi con RDF.

[24] I metadati Dublin Core <http://dublincore.org/> comprendono *Title*, *Creator*, *Subject*, *Description*, *Publisher*, *Contributor*, *Date*, *Type*, *Format*, *Identifier*, *Source*, *Language*, *Relation*, *Coverage* e *Rights*.
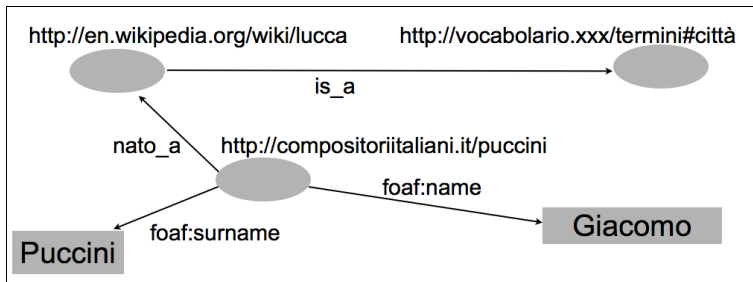
Figura 6. Grafi in RDF

c) Per esplicitare in modo formale un determinato campo di conoscenze, è necessario un ulteriore passaggio che permetta di associare i concetti a regole logiche d'uso. Questo compito è assegnato alle ontologie. In filosofia, un'ontologia è una teoria sulla natura dell'essere di tipi di oggetti; in senso più tecnico, è un accordo che si basa su definizioni condivise e che rappresenta la base concettuale su cui operiamo. Il termine è diventato di uso comune tra i ricercatori che si occupano di intelligenza artificiale e del Web, e in questo significato l'ontologia è un documento che definisce in modo formale le relazioni tra termini[25]. Nel grafo della figura sopra è già presente un'ontologia. Si tratta del link "is_a" tra "lucca" e "città". "is_a" è in effetti una relazione che fa parte dei mattoni per creare un'ontologia. Relazioni come "is_a" sono in un certo senso predefinite: sono gli assiomi del linguaggio, i connettori logici.

La tipica ontologia per il Web ha una tassonomia e un insieme di regole di inferenza. La tassonomia definisce classi di oggetti e le relazioni tra essi. Così, il significato dei termini in una pagina web può essere stabilito da puntatori che linkano a un'ontologia. Un agente semantico che cerchi l'ospedale di Pisa e trovi "Santa Chiara" deve poter capire che non cerchiamo una biografia di Santa Chiara, né una chiesa o una Chiara qualunque. L'ontologia è un albero che permette di esprimere restrizioni sui termini (condizioni), quindi relazioni. Attraverso le relazioni della logica descrittiva si può esprimere in maniera formale qualsiasi oggetto o concetto. Le ontologie dunque contengono le specifiche dei concetti necessari a comprendere un dominio di conoscenza, il vocabolario corri-

[25] Nel campo delle "*computational ontologies*", si parla di ontologie con due accezioni diverse, entrambe corrette. La prima accezione intende con ontologia l'insieme sia dei dati, sia dei vocabolari; nella seconda accezione, l'ontologia è l'insieme dei soli vocabolari. In logica descrittiva si fa distinzione tra Abox e Tbox come due componenti distinti delle ontologie. La Tbox definisce un vocabolario e le relazioni tra gli elementi del vocabolario, la Abox è un insieme di asserzioni compatibili con quanto definito dalla Tbox. Qui intendiamo l'ontologia nella seconda accezione.

spondente, e il modo in cui concetti e vocabolari sono collegati e in cui sono definite e descritte le classi, le istanze e le loro proprietà[26].

Un'ontologia può essere formale o informale. Il vantaggio di un'ontologia formalizzata risiede nel fatto che questa è *machine-readable,* vale a dire che un computer può compiere dei ragionamenti a partire da essa. Di contro, lo svantaggio sta nella difficoltà di implementare tali costrutti formali.

È intuitivo osservare che due basi di dati possono usare identificatori differenti per il medesimo concetto. Un programma che voglia confrontare o combinare informazioni nei due database deve sapere che più termini possono riferirsi al medesimo oggetto. Idealmente, il programma deve essere in grado di riconoscere due termini come sinonimi. La soluzione a questo problema è data dalle altre ontologie[27] – le quali debbono mettere in correlazione i dati usando una lingua franca in modo da facilitarne la condivisione. Si veda nella figura sotto un esempio.

Web dei dati

ArteAlucca.com

Opera	Titolo	Luogo	Città
9328	Ilaria del carretto	Duomo di San Martino	Lucca

chieseditalia.com

ID	Nome	Località	Data nascita
2435	Duomo di San Martino	Lucca	22/12/1858

chieseditalia.com

ID	Nome	Località	Data nascita
2435	Duomo di San Martino	Lucca	22/12/1858

geografiaItalia.com

CAP	Nome	Long	Lat
55100	Lucca	3249784	34322

compositori.com

ID	Nome	Luogo Nascita	Data nascita
45	Giacomo Puccini	Lucca	22/12/1858

Figura 7. Esempio di incrocio di dati da basi di dati distinte.

[26] T. Gruber, *A translation approach to formal ontologies*, «Knowledge Acquisition», 5, 25, 1993, pp. 199–200, <http://ksl-web.stanford.edu/KSLAbstracts/KSL-92-71.html>.

[27] Le ontologie sono scritte in RDF Schema, *Onthology Web Language* (OWL), e in altri linguaggi appositi, ad esempio DAML, DAMOIL.

Sul piano dell'implementazione, l'ontologia esistente più diffusa è Wordnet, che descrive i concetti (sinonimi, contrari, relazioni tra concetti) in diverse lingue. La peculiarità di tale sistema è che è in grado di effettuare ricerche di "dipendenza concettuale", come ad esempio le relazioni meronimiche (parte-tutto), sineddotiche o metonimiche (per esempio, se cerco un ospedale nella città di Lucca, è essenziale che il sistema sappia in qualche modo che Lucca è in Toscana e che la Toscana si trova in Italia).

Si osservi che chiunque può creare un propria ontologia di riferimento (sperando che venga condivisa) o adottarne una esistente, optando così per una altrui prospettiva. In genere, le ontologie sono create all'interno di comunità di conoscenza. Un problema, riscontrato e segnalato in particolare dai filosofi, è che ciascuno definisce il mondo a suo piacimento, e che proprio la creazione di queste definizioni è parte integrante del mestiere del ricercatore in scienze umane e sociali. Dunque, esiste un alto rischio di creare sistemi che si sovrappongono o che sono in conflitto, mentre la ricerca di un'unica ontologia, se da una parte sarebbe auspicabile, può comportare il pericolo insito nella rigidità delle interpretazioni di fatti e fenomeni.

Tuttavia, tale limite o pericolo può diventare una ricchezza. È infatti auspicabile che vengano create molte e diverse ontologie, poiché più sistemi di classificazione sono in grado di stabilire una più corrispondenze tra nomi e oggetti di quanto non possa fare un unico vocabolario; inoltre, è possibile creare meta-ontologie che stabiliscano corrispondenze tra sinonimi, e che mettano a loro volta in relazione vocabolari distinti. Ontologie diverse possono essere intersecate in modo da stabilire analogie, cioè relazioni di equivalenza. Spesso, gruppi diversi elaborano indipendentemente concetti molto simili e descrivere la relazione che esiste tra questi porta grandi benefici. Per un software, potrà essere dunque normale avere a che fare con ontologie differenti e magari in parte conflittuali nello stesso dominio di conoscenza, e sarà un compito del programma stesso mostrarne le diversità, aiutando così a migliorare la comprensione del concetto in questione.

Un'altra garanzia di pluralità, per la quale i teorici del web semantico si distinguono dalle teorie dell'intelligenza artificiale, è data dalle implicazioni della cosiddetta "*Open World Assumption*", secondo la quale il valore di verità di un'asserzione è indipendente dalle conoscenze dell'osservatore. Per fare un esempio, se la mia asserzione è "Francesca" "è" "italiana", la risposta alla domanda: Paolo è italiano? Sarà: non lo so. I linguaggi RDF e OWL si basano su questo assunto della logica formale. L'assenza di una particolare affermazione sul web significa semplicemente, in principio, che non è stata ancora esplicitata, al di là del fatto che possa essere vera o falsa e dalle credenze dell'osservatore in merito. In altre parole, questo significa che se un osservatore non sa se un'asserzione è vera non può inferire che essa sia falsa (come invece accadrebbe in un mondo in cui vale la "*Closed World Assumption*").

L'unione dei dati espressi in RDF e dei dati espressi in un linguaggio delle ontologie (ad esempio OWL) permette di inferire nuova conoscenza. È questo lo strato della logica, in cui trova spazio il "*reasoning*", l'inferenza di nuova conoscenza: per costruire procedure analoghe al ragionamento software appositi devono essere in grado di collegare i termini, una funzione che sarà resa possibile tramite i linguaggi di inferenza. I linguaggi di inferenza consentono alle macchine di convertire dati da un formato all'altro, riconoscendo due termini come identici e traducendoli, un po' come fa un dizionario bilingue. Si tratta di una funzione fondamentale poiché nessuno ha il potere di definire un termine per tutti; questi linguaggi si occupano pertanto di identificare relazioni tra basi di dati per stabilire la presenza di sinonimi. Per tornare all'esempio di Lucy e Pete alla ricerca di un medico per il trattamento prescritto alla madre, grazie ai linguaggi di inferenza potrà essere trovato l'ospedale giusto anche usando i termini "clinica", "casa di cura" o "hospital", e via dicendo. Questa specifica funzione è assegnata a motori di ricerca logici, motori in grado di applicare le regole della logica per stabilire se le risposte ottenute in una ricerca iniziale sono utili o no. Nell'esempio, è un motore che riesca a rintracciare le informazioni relative alla loro specifica richiesta e a combinare i criteri da loro definiti: il trattamento richiesto e la copertura assicurativa dell'assistita, la distanza dalle rispettive abitazioni, gli orari in cui si sono resi disponibili, e altre informazioni rilevanti.

d) Infine, un concetto essenziale è quello di fiducia (*Trust*).

> La rete di fiducia è un modello essenziale del modo in cui lavoriamo realmente come persone. Ognuno di noi costruisce la sua rete sin dall'infanzia. Man mano che decidiamo cosa linkare, cosa leggere o comprare sul web, un elemento che entra a far parte della nostra decisione è quanto possiamo fidarci dell'informazione che vediamo. Potremo fidarci del nome dell'editore, delle pratiche di tutela della privacy, delle motivazioni politiche? Certe volte impariamo nel modo peggiore che non dobbiamo fidarci, ma più spesso ereditiamo la fiducia da altri, da un amico o da un insegnante o da un familiare, o da raccomandazioni edite oppure da garanzie di terzi come la banca o il dottore.[28]

La costruzione di reti di fiducia è necessaria a implementare sistemi di filtro dell'informazione secondo criteri di qualità condivisi ma soggettivi, ed è uno dei temi su cui discute da anni il W3C[29]. In particolare, sul Web il concetto di fiducia assume notevole importanza da un punto di vista socio-culturale, sul piano cioè della trasformazione dei comportamenti dei navigatori della rete. La qualità e la rilevanza di una pagina dipendono da come viene percepita da parte di un lettore, elementi che di-

[28] T. Berners-Lee, *L'architettura del nuovo Web*, cit., pp. 137-38.
[29] *Ivi*, pp. 134-37.

pendono da una serie di credenze soggettive. Sono queste credenze soggettive a rientrare nella nozione di fiducia (*Trust*). La fiducia è un importante fattore nello sviluppo del Web, in particolare per quanto riguarda l'uso del Web come mezzo di comunicazione scientifica, visto che l'architettura del Web, e in particolare l'assenza di un'autorità centralizzata, consente l'anonimato e non impedisce la copia. Tuttavia, studiare come si costruisce la fiducia online è particolarmente difficile a causa di diversi fattori: i molteplici contesti in cui si svolgono le interazioni; il fatto che la fiducia in un sistema e in un individuo vengano assimilate; infine, gli studi empirici, per quanto rigorosi e ben condotti, presentano definizioni diverse del termine, rendendo pressoché impossibile una comparazione. Inoltre la fiducia non è un fenomeno statico ma dinamico: spesso la fiducia in un sito viene costruita nell'arco del tempo, rendendo difficile quantificarla nel breve periodo; abbiamo visto che l'*Impact Factor* si calcola su un arco di tempo di due anni, ma sul Web due anni sono un periodo breve, e del tutto arbitrario. Infine, utenti con diversi livelli di esperienza esprimono livelli di fiducia tipicamente differenti. Tutto ciò implica che la fiducia non possa essere prodotta semplicemente creando lo strumento e la tecnologia giusta. Ad esempio, qual è il giusto compromesso fra un sistema molto efficiente e costoso, e uno più scarno e meno efficiente, ma più economico e facile da usare? È necessario riflettere su questi aspetti da un punto di vista tecnico, ma anche filosofico e politico.

Su un piano individuale, si tende ad affidarsi a meccanismi di *feedback* sul comportamento, tramite cui un utente acquisisce una certa reputazione. La reputazione è un elemento chiave per la fiducia.

Per capire come aggregare e come valutare le reputazioni, bisogna porre alcune questioni preliminari: «Quali metodi renderanno possibile la raccolta di *rating* che definiscano l'affidabilità di un utente Web? Come dovrebbero essere aggregati tali *rating*? Come dovremmo ragionare su queste aggregazioni? E in che modo queste dovrebbero essere rese pubbliche?»[30]

Il sistema di eBay, in cui i *rating* sono valori +1 o −1 a cui si aggiungono annotazioni testuali è l'esempio più noto di costruzione della reputazione. Nonostante la sua affidabilità non sia inquestionabile, il suo successo dimostra, nei fatti, l'efficacia di tale meccanismo. Tuttavia, in campo scientifico tale sistema si rivelerebbe in qualche misura povero e inefficace, in quanto il campo scientifico è composto da moltissime comunità piccole (nicchie), e sui piccoli numeri le distorsioni introdotte da giudizi poco motivati o ingiusti sarebbero difficilmente ovviabili. Le diverse dimensioni delle comunità richiedono pertanto di pensare sistemi differenti.

[30] Cfr. T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, *A Framework for Web Science*, cit., p. 89, e in generale, si vedano le pp. 88-94.

Una questione complessa, collegata alla precedente, è quella di trovare metriche adeguate a misurare la fiducia per i *rating* individuali, e algoritmi adeguati a creare aggregazioni. La valutazione delle metriche della fiducia è inevitabilmente difficile anche se non impossibile a patto che esista un contesto sufficiente a rendere la valutazione significativa. In alcuni casi, le metriche quantitative sono dunque inappropriate. Per esempio, nel valutare una fonte di informazione, può accadere che l'utente abbia bisogno di confrontare le annotazioni e le analisi degli altri, più che affidarsi a un voto, o a una media di voti. Tuttavia, in campi compatti e limitati, il *mark-up* semantico dei documenti potrebbe rivelarsi un ottimo metodo per valutare l'affidabilità e la qualità di un documento. Resta comunque il fatto che un ampio numero di giudizi di esperti, indipendentemente dal modo in cui sono raccolti, rappresentati o aggregati, è essenziale a sviluppare un sistema di riconoscimento di fiducia durevole e affidabile. L'architettura del Web, e la facilità con cui è possibile linkare nuova informazione a quella esistente, permette la costruzione della fiducia in modo semplice. Applicazioni che sfruttano tali potenzialità stanno cominciando a emergere; tra queste, vale la pena ricordare FOAF[31], acronimo di Friend Of A Friend, un'ontologia basata su RDF/OWL che descrive le relazioni di un individuo con i suoi amici.

Capire come informazioni sulla fiducia, la credibilità, l'attendibilità e la qualità di un documento possano essere estratte dalla struttura del Web è un tema centrale nello studio dei *social network*. Ricerche su questo tema potrebbero anche essere usate come metodo per fare una classifica delle pagine nelle ricerche sul web. Le tecniche di generazione di fiducia potrebbero avere sul Web semantico un ruolo simile a quello che algoritmi come PageRank hanno sul Web tradizionale. Un campo essenziale per la valutazione della ricerca scientifica, come vedremo nel prossimo capitolo.

La rete semantica è dunque in grado di descrivere l'informazione, poi di dedurre nuova conoscenza e infine di ragionare a partire da essa. È questo, in sintesi, il sistema necessario a far funzionare i nostri agenti semantici nella storia di Lucy e Pete.

> Quando migliaia di moduli saranno collegati in tutto il campo "cognomi", allora tutti quelli che analizzeranno il Web capiranno che è un importante concetto comune. La cosa bella è che nessuno dovrà compiere realmente questa analisi. Il concetto di "cognome" comincerà semplicemente a emergere come caratteristica importante di una persona. Come un bambino che impara un'idea tramite contatti ripetuti, la Rete Semantica "impara" un concetto tramite contributi ripetuti da diverse fonti indipendenti. […] Il ragionamento che sta dietro questo approccio, quindi, è che non esiste un magazzino centrale del-

[31] <http://www.foaf-project.org/>.

> l'informazione, e nessuna autorità su alcunché. Collegando le cose tra di loro potremo fare molta strada verso la creazione di una comprensione comune. La Rete Semantica funzionerà quando ci saremo messi d'accordo sui termini, ma anche se non ci saremo riusciti.[32]

La pretesa della rete semantica non è dunque quella di poter rappresentare tutti i dati o il sapere in qualche ristretto insieme di formalismi, ma piuttosto fare sì che la possibilità di linkare i dati a nuovi dati permetta di usarli in modo sempre più ampio, confidando nell'intelligenza distribuita e interconnessa dei navigatori. L'ambizione del W3C consiste nell'incrementare i dati a disposizione, e nel valorizzarli tramite nuove tecnologie che si aggiungono ai pilastri del Web. L'estensione del livello di inferenza che può essere ottenuta automaticamente non è dunque uno scopo ma semmai una conseguenza auspicabile[33].

Questa evoluzione facilita la creazione di gruppi con interessi comuni, vere e proprie comunità aperte in rete. Nonostante le applicazioni del Web Semantico siano per il momento confinate per lo più all'ambito accademico, le loro possibili implicazioni cominciano a diventare evidenti. Restando nei confini della ricerca accademica, non è difficile vedere le implicazioni di tale rivoluzione non solo nella scienza del Web, ma anche nella pratica della ricerca in generale. Grazie alla costruzione di un Web ipertestuale, un gruppo di qualsiasi dimensione può comunicare con facilità, acquisire sapere e veicolarlo velocemente, superare le incomprensioni e ridurre la ridondanza degli sforzi. «L'universalità, scrive ancora Tim Berners-Lee, deve esistere in molte dimensioni. Tanto per cominciare, dobbiamo essere in grado di collegare tra loro molti documenti, dalle bozze alle successive versioni fino ai documenti definitivi»[34]. Un passaggio essenziale a non perdere le tappe del processo di un ragionamento, perché quando facciamo ricerca il processo tramite cui si raggiunge un risultato è importante al pari del risultato stesso.

> Qualora nuove persone entrassero in un gruppo, avrebbero a disposizione tutto un passato di decisioni e motivazioni. Quando lasceranno il gruppo, il loro lavoro sarà già stato assorbito e integrato. E come interessante bonus, l'analisi automatica della rete di conoscenza potrebbe consentire ai partecipanti di trarre conclusioni sulla gestione e l'organizzazione della loro attività collettiva, un'impresa che non sarebbe stata possibile altrimenti[35].

[32] T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt, D.J. Weitzner, *A Framework for Web Science*, cit., p. 163.

[33] *Ivi*, p. 30.

[34] *Ivi,* p. 128.

[35] *Ivi,* p. 143.

Uno scenario che non è molto distante da quello prefigurato, nel 1945, da Vannevar Bush.

3. Ragnatele sociali: dal Web 2.0 al Web 3.0

Di pari passo con la definizione e lo sviluppo del Web semantico, si è assistito a un'evoluzione sociale del Web che, ad un primo sguardo, sembra divergere dalla prospettiva abbracciata da Berners-Lee e dagli architetti del Web che fanno capo al W3C. Dal 2004 si è infatti cominciato a parlare di "Web 2.0", termine coniato dall'editore americano Tim O'Reilly per definire «un insieme di tendenze economiche, sociali e tecnologiche che formano collettivamente la base per la futura generazione di Internet, che diventa un mezzo più maturo» e definito sulla base di alcune particolari caratteristiche: partecipazione degli utenti, apertura ed effetti di rete[36]. Questa nuova definizione è stata apertamente criticata dallo stesso Berners-Lee, che ha affermato che l'attuale evoluzione del Web non è altro che «l'uso di standard creati dalle stesse persone che lavorano sul cosiddetto Web1.0» (in primo luogo URL/URI, (X)HTML, HTTP), uno spazio di documentazione progettato dal principio come «spazio collaborativo, in cui le persone potessero interagire»[37].

È certamente vero che l'aspetto tecnologico del Web 2.0 è funzionale alle caratteristiche delle reti sociali e della cooperazione tra utenti, e le nuove tecnologie che vengono sviluppate poggiano sui princìpi architettonici di Internet, Web e web semantico, vale a dire l'apertura e l'universalità che si esplicano nell'architettura a strati, nella decentralizzazione, nella "*open world assumption*", nella centralità degli URI e nel principio del "*data linking*".

Tuttavia, il termine Web 2.0 si è diffuso divenendo l'espressione con cui si sintetizzano i più recenti fenomeni tecnico-sociali che trovano spazio sul WWW. Si tratta di siti come MySpace e Facebook, reti sociali tra persone che condividono gli stessi interessi (il numero di utenti di Facebook ha raggiunto, nell'aprile 2009, i duecento milioni), il cui successo è dato dalla facilità con cui è possibile trovare e collegarsi ad altri utenti già presenti nel "mondo piccolo" del *network*; ma anche come YouTube, eBay, e le migliaia di siti dalle caratteristiche analoghe, basati su un modello di Web inteso come canale essenziale per reperire, creare, gestire e

[36] Cfr. T. O'Reilly, *Che cos'è web 2.0*, 2004, <http://www.awaredesign.eu/articles/14-Cos-Web-2-0>. Si veda anche il video, molto istruttivo, *Web 2.0 ... The Machine is Us/ing Us* di Michael Wesch, docente di antropologia culturale alla Kansas State University, online su Youtube all'URL <http://www.youtube.com/watch?v=6gmP4nk0EOE>.

[37] Cfr. l'intervista di Scott Laningham a Tim Berners-Lee del 22 agosto 2006 all'URL: <http://www.ibm.com/developerworks/podcast/dwi/cm-int082206txt.html>.

condividere informazioni, tramite aggregatori dell'accesso a servizi che, fino a pochi anni fa, erano disponibili solo off-line.

Un principio fondamentale di questa evoluzione è il cosiddetto "*mash up*" (letteralmente: "poltiglia")[38], che si traduce in applicazioni on-line che combinano o miscelano «dati e servizi provenienti da due o più fonti per creare qualcosa di nuovo»[39]. Un esempio del principio è la trasmissione di RAI3 *Blob*, che rimescola spezzoni di programmi televisivi e di film accompagnati da nuovi titoli e intestazioni che attraverso la somma dei frammenti, restituiscono un nuovo significato. Un altro esempio di *mash up*, questa volta sul Web, è quello della combinazione di Google Maps con altri servizi, come flickr, un sito che consente agli utenti di pubblicare e condividere fotografie sul Web; nello specifico, la combinazione di Google Maps e flickr permette di localizzare e visualizzare le immagini di flickr su una mappa[40].

Da un punto di vista dello sviluppo delle tecnologie, gli utenti sono coinvolti nel processo di programmazione e trattati come co-sviluppatori, il software viene rilasciato spesso e la sua pubblicazione non è subordinata all'accumularsi di miglioramenti e innovazioni sostanziali, e il modello di progettazione che viene seguito è indipendente dall'hardware (ora sempre più aperto anche ai cellulari) e a strati.

[38] In termini informatici, indica un'applicazione che usa contenuto da più sorgenti per creare un servizio completamente nuovo. Cfr. <http://it.wikipedia.org/wiki/Mash-up_(informatica)> (10/09).

[39] J. Musser, *Web Service e mashup*, in V. Di Bari (a cura di), *Web 2.0*, Il Sole 24 ore, Milano 2007, p. 115.

[40] Cfr. <http://www.earthalbum.com/>; <http://loc.alize.us/>. La tecnologia che dà facoltà di farlo sono le API *(Application Programming Interface*, <http://it.wikipedia.org/wiki/Application_programming_interface>) insiemi di procedure, strutture dati, classi di oggetti e protocolli forniti dalle librerie e dai servizi del sistema operativo per supportare l'implementazione di applicazioni, che, quando liberamente disponibili, si dicono aperte (*Open* API). «Nel mondo del software l'attenzione intorno ai *mash up* viene sostenuta dalle possibilità di uno sviluppo più rapido e agile, basato su dati e servizi riutilizzabili su larga scala. Lo stesso HousingMaps.com, che è stato creato da un solo programmatore nell'arco di pochi giorni, è la dimostrazione di queste possibilità. La realizzazione di un'applicazione tanto sofisticata in così poco tempo sarebbe stata impensabile solo pochi anni fa. Tuttavia, con i preziosi sistemi per la generazione di contenuti e gli strumenti ricchi e potenti che si trovano su Google Maps, tutto è stato reso più semplice.

ChicagoCrime.org è un altro importante *mash up* basato sulle mappe. Riutilizza una serie di dati sulla criminalità di per sé di difficile lettura e li rende molto più comprensibili grazie a una migliore disposizione, presentazione e mappatura. Nel mercato del commercio elettronico, quasi la metà di tutti i beni in lista su eBay vengono presentati tramite strumenti basati sulle API dei servizi Web di eBay. Oggi Amazon dispone di oltre 250.000 programmatori registrati che creano migliaia di applicazioni. Nel terzo quadrimestre del 2006 più del 50% di tutte le transazioni su Salesforce.com sono state caricate grazie ad applicazioni esterne utilizzando API specifiche e non tramite Web standard della stessa Salesforce.com». J. Musser, *Web Service e mash up*, cit., p. 116.

Il Web fa sempre più da specchio alla società umana di cui riflette gli interessi, che si estendono su un arco molto ampio di assunti, valori e culture. La rete è usata per condividere informazioni, per divertirsi, per vendere e comprare, e per fare scienza in modi radicalmente nuovi. Il Web 2.0 è stato così rappresentato «come un insieme di princìpi e di procedure che collegano un autentico sistema solare di siti che fanno propri tali principi, in tutto o in parte, a una distanza variabile da tale centro»[41].

Ma quali sono i principi che caratterizzano la fisionomia del Web 2.0? I tentativi di sintesi del fenomeno sono molti e variano tra loro. Tuttavia, è possibile rintracciare sei concetti, o architravi portanti della sua struttura tecnica e sociale:

1) "Il web come piattaforma"

I servizi web sostituiscono sempre più le applicazioni desktop. Un esempio è l'account di Google, che offre gratuitamente agli utenti connessi l'integrazione di servizi prima disponibili solo in locale, come ad esempio l'email (gmail), editor di scrittura tipo Word e suite di applicazioni come Microsoft Office (Google docs), il calendario (Google calendar) e moltissimi altri[42]. Google diviene così l'intermediario tra utenti e informazione, creando veri e propri *desktop* online in cui le applicazioni non sono concepite come pacchetti in vendita ma fornite come servizio, senza licenze o condizioni di vendita, e senza richiedere la portabilità su piattaforme diverse affinché i clienti possano utilizzare il software sulle proprie macchine[43]. Un corollario di questa trasformazione è che il software è concepito sempre più come un servizio e non come un prodotto. Il modello Google si contrappone così al modello tradizionale *à la* Microsoft:

> da una parte, un singolo fornitore di software, la cui massiccia base e il sistema operativo strettamente integrato con le API permettono di controllare il paradigma di programmazione; dall'altra, un sistema senza un proprietario, tenuto insieme da una serie di protocolli, standard aperti e accordi di cooperazione[44].

Con il Web 2.0 si supera il concetto di sito Web come entità autonoma incapace di comunicare con il mondo esterno; si diffondono molte applicazioni sviluppate dal basso da comunità decentralizzate, che rientrano nel cosiddetto *social software*; inoltre, i dati online diventano indi-

[41] T. O'Reilly, *Che cos'è web 2.0*, cit.

[42] Per usufruirne, è sufficiente avere un account google, e accedere a tali servizi dalla pagina<http://www.google.it/intl/it/options/> (in italiano).

[43] In termini tecnici, sono servizi "platform independent", vale a dire indipendenti dal sistema operativo dell'utente.

[44] T. O'Reilly, *Che cos'è web 2.0*, cit.

pendenti dalle applicazioni e possono essere utilizzati da servizi diversi. Gli esempi di siti di questo tipo sono molteplici: si vedano, tra gli altri, Librarything e Anobii[45], nati per creare cataloghi di libri, oltre a schede di lettura e commenti, e per socializzare, condividere e scambiarsi bibliografie e Citeulike e Connotea[46], che hanno lo stesso scopo dei due siti precedenti ma sono finalizzati alla creazione di bibliografie scientifiche. Gli ultimi due in particolare mirano a sostituire programmi come EndNote[47], un software proprietario e a pagamento che dev'essere installato sul computer di chi lo usa.

2) Centralità dei dati

«Senza i dati, scrive ancora O'Reilly parafrasando una nota affermazione di Kant, gli strumenti sono inutili; senza il software, i dati sono ingestibili». Google si configura così come un insieme di strumenti software e un ampio database specializzato[48]. La disponibilità di un'enorme quantità di dati è una caratteristica fondamentale del nuovo web. Le figure 8 e 9 nella pagina seguente mostrano sorgenti dati semantiche sul Web, rispettivamente nel 2007 e nel 2009.

3) Contenuti prodotti dagli utenti (User generated content)

Un importante corollario del principio precedente è che i contenuti sono generati dagli utenti e condivisi all'interno di comunità. In questo modo, sono gli stessi utenti a creare il "servizio"[49]. È questo un elemento che caratterizza il Web 2.0 e il *social software*:

> Da *"The computer moves in"* a *"Yes, you. You control the information age. Welcome to your world"*. Dall'ormai celebre copertina del *Time* del 3 gennaio 1983[50], che nominava il personal computer come *"machine of the year"*, a quella, quasi un quarto di secolo dopo, del 25 dicembre 2006[51], in cui il titolo di *"man of the year"* va alla seconda persona, sia singolare sia plurale, della lingua inglese, scritta, o meglio, digitata al centro di un monitor.

"La persona dell'anno", recita il Time, "sei tu". "Benvenuto nel tuo mondo".

45 < http://www.librarything.it/>; <http://www.anobii.com/>.

46 <http://www.citeulike.org>; <http://www.connotea.com>.

47 <http://www.endnote.com/>.

48 T. O'Reilly, *Che cos'è web 2.0*, cit.

49 Il Web è ricco di esempi. Tra essi, oltre al già citato Flickr <http://www.flickr.-com> vale la pena menzionare Youtube e ebay, che non hanno bisogno di spiegazioni.

50 <http://www.time.com/time/covers/0,16641,19830103,00.html>.

51 <http://www.time.com/time/covers/0,16641,20061225,00.html>.

Figura 8. Sorgenti dati semantiche sul Web, 2007

Figura 9. L'immagine sopra, aggiornata al marzo 2009[52]

52 Da <http://linkeddata.org/>.

> Se nel gennaio del 1983 la figura umana veniva presentata stilizzata e glaciale di fronte al computer, sul cui schermo comparivano numeri e grafici, in un ambiente cupo dove, per contrasto, solo la sedia ed il tavolo erano caratterizzati da un aspetto tradizionale e da colori caldi[53], nella copertina del 2006 questa presenza è totalmente astratta, ridotta all'essenza di un pronome personale, al posto dei grafici nel monitor, ora a cristalli liquidi.[54]

La rete di calcolatori lascia così spazio alle reti sociali sul web, in cui gli utenti producono e condividono contenuti, arricchendoli di ulteriori dati e collegandoli tra loro. Nasce un nuovo modo di concepire la comunicazione: chiunque abbia accesso a Internet crea contenuti e li condivide con altri; chiunque può replicare a contenuti proposti, classificarli in modi nuovi e votarli, dando vita a un'agorà virtuale in cui l'interattività è un elemento fondante.

I blog sono, in questo contesto, un importante strumento di pubblicazione interamente nelle mani degli utenti.

> La "blogosfera" può essere considerata un nuovo equivalente peer-to-peer di Usenet e dei bulletin-board, le aree conversazionali del primo internet. Non solo è possibile iscriversi ai siti degli altri e collegarsi facilmente ai singoli commenti su una pagina: tramite un meccanismo noto come trackback si può anche vedere quando qualcun altro si collega alle proprie pagine e rispondere, con link reciproci o aggiungendo commenti[55].

I "*trackback*", link monodirezionali che creano l'effetto di collegamenti a due vie, incrementano il grado di connessione dei contenuti e creano zone in cui il Web si presenta come una rete indiretta e più facile da navigare. Gli effetti sull'editoria online di questa funzione sono per certi versi rivoluzionari:

> L'avvento degli *user generated content* sta impattando soprattutto sulla riorganizzazione dello scenario dell'informazione e dell'intrattenimento (il settore *media & entertainement*). Chi produceva (e traeva profitto) dalla realizzazione di contenuti (i cosiddetti *content provider*) ha dovuto (o dovrà, velocemente) riposizionarsi perché nuovi *player* sono subentrati e hanno assunto in breve tempo posizioni rilevanti. Alcuni *content provider* tradizionali hanno già integrato alla

[53] Un approccio visuale simile sarà utilizzato l'anno successivo nel famoso spot pubblicitario "1984" di Ridley Scott per l'Apple Macintosh. Il video dello spot è disponibile online a diversi indirizzi, per maggiori informazioni vedi <http://en.wikipedia.org/wiki/1984_%28television_commercial%29>.

[54] F. Meschini, *eContent: tradizionale, semantico o 2.0?*, 2006, <http://dspace.unitus.it/handle/2067/162>.

[55] T. O'Reilly, *Che cos'è web 2.0*, cit.

> creazione di contenuti tradizionali la creazione di ambienti e strumenti che permettano l'accesso a contenuti pubblicati da non professionisti, come per esempio la BBC con i suoi innovativi progetti interattivi[56].

Il sito che meglio esemplifica questo principio è YouTube[57], progetto creato nel 2005 e acquistato nel 2006 da Google, che consente di condividere video creati dagli utenti e raccoglie centinaia di migliaia di filmati in tutte le lingue.

Un elemento che ha arricchito questa tipologia di siti web sono le "folksonomie", neologismo creato nel 2003 da Thomas Vander Wal che deriva dalla contrazione dei termini "*folks*" e "*taxonomy*", per indicare stili di categorizzazione collaborativa che utilizzano parole chiave liberamente scelte, generalmente definite *tag*. «In maniera più semplice e concreta, questo termine si riferisce alla metodologia utilizzata da gruppi di persone che collaborano spontaneamente per organizzare in categorie le informazioni disponibili attraverso Internet»[58]. Se l'uso di ontologie aggiunge una struttura ai dati, con le folksonomie la struttura emerge in modo organico dall'organizzazione delle richieste informative dei singoli individui. Le folksonomie sorgono quando un ampio numero di persone è interessato a qualche informazione e incoraggiato a descriverla – in gergo, a "taggarla". Si tratta dunque di sistemi di classificazione distribuiti, ed è questo elemento a costituirne la caratteristica principale[59].

All'aumentare dei *tag* inseriti dagli utenti, questi tendono a essere riusati e applicati a nuovi elementi da altri utenti.

> Il valore aggiunto delle folksonomie sta proprio nell'aggregazione e nella condivisione, cioè nella natura sociale del fenomeno: la loro forza non è la precisione ma l'ampia adesione popolare: la gente, appunto. Anziché sforzarsi di creare sistemi di classificazione che incontrino i modelli mentali degli utenti, si fa in modo che siano gli utenti stessi – dal basso – a far emergere spontaneamente (attraverso un processo collaborativo e una sorta di selezione naturale) modelli mentali comuni, condivisi. Sarebbe infatti più giusto definire le folksonomie come sistemi di classificazione distribuita[60].

Certamente, la libertà nell'assegnare parole chiave a un contenuto può creare un eccesso di informazione e una certa confusione. Tuttavia, i

[56] V. Di Bari, *Introduzione al Web2.0*, in Id. (a cura di), *Web 2.0*, cit., p. 18.

[57] <http://www.youtube.com>.

[58] <http://it.wikipedia.org/wiki/Folksonomia> (10/09).

[59] Cfr. E. Quintarelli, *Folksonomies: power to the people*, «ISKO Italy-UniMIB meeting», 2005, <http://www.iskoi.org/doc/folksonomies.htm>.

[60] L. Rosati, *La ricerca si fa sociale: il tagging e la folksonomia*, in V. Di Bari (a cura di), *Web 2.0*, cit., p. 66.

tag sono generati dalle interazioni che avvengono nel mondo reale e rivelano connessioni reali tra contenuti e utenti. Tali strutture permettono alla semantica di emergere da accordi impliciti (*ex post*), diversamente da quanto avviene nella costruzione di ontologie, le quali indicano accordi espliciti (*ex ante*); il campo delle dinamiche semiotiche ha come premessa l'idea che gli accordi comunicativi o i sistemi di organizzazione dell'informazione spesso si sviluppino attraverso processi decentralizzati di invenzione e negoziazione[61].

La distinzione tra ontologie e folksonomie ha un notevole rilievo su un piano filosofico, e merita qualche parola in più. È stato osservato che le folksonomie sono preferibili all'utilizzo di ontologie controllate e centralizzate. Se infatti la possibilità di annotare pagine web usando vocabolari controllati può accrescere le possibilità di trovare ciò che cerchiamo con esattezza, d'altra parte una base ampia ed eterogenea di utenti non consente che siano molte le persone che scelgono di adottare o di mantenere un'ontologia complessa. Creare ontologie richiede un investimento di tempo e di energie molto elevato. Tuttavia è altrettanto vero che le parole chiave scelte da una persona in modo indipendente e scoordinato dagli altri possono essere inutili o inaccurate, e anche se l'uso di folksonomie è assai diffuso, queste rimangono uno strumento di classificazione impreciso. Allo stesso tempo, però, la diffusione di strumenti per l'uso di ontologie è assai più lenta di quanto non stia avvenendo per le applicazioni sociali che si servono delle folksonomie. Dunque, è inopportuno e improprio considerare folksonomie e ontologie come strategie mutuamente escludentisi – molto meglio invece adottarle come complementari. Le prime si rivelano così di grande utilità per la ricerca di documenti da parte degli utenti; le seconde sono essenziali per il recupero automatizzato di dati, in modo particolare in campo scientifico[62].

> Spesso la folksonomia viene vista come un'ontologia debole, dotata di una struttura leggera ma di valore (approssimativamente circoscritta all'oggetto che viene etichettato, ai tag e all'identità della persona che effettua l'operazione di etichettatura). Credo fermamente – scrive Vander Wal – in una folksonomia i cui risultati saranno migliorati grazie all'abbinamento con la tassonomia. Le aree dove la tassonomia è debole (dove non riesce a far emergere i risultati con facilità, non è esaustiva, non è facile da applicare e richiede tante risorse per essere creata e gestita) sono le aree in cui la folksonomia è veramente forte. Al contrario, le aree dove la folksonomia è più debole (disordine, scarsa rintracciabilità e lentezza a far emergere i dati) sono ambiti nei quali la tassonomia può esserci d'aiuto fornendoci un qualche soste-

[61] T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, *A Framework for Web Science*, cit. p. 32.

[62] *Ivi*.

gno. Oggi non sono disponibili strumenti che consentono questo mix, ma sono sempre più richiesti dalle imprese[63].

In realtà, gli esempi di convergenza tra l'impostazione del web semantico e quella del Web 2.0 non mancano. Tra gli altri, vale la pena ricordare *Visual Thesaurus*[64], un dizionario della lingua inglese, che utilizza il software di visualizzazione *ThinkMap*[65] e che è basato su *WordNet*[66], un database lessicale della lingua inglese il cui progetto è stato avviato nel 1985, in cui i termini sono organizzati tra di loro sulla base di rapporti semantici che caratterizzano le classificazioni ontologiche (iponimia, ipernimia, olonomia e meronimia). Il vocabolario presenta i vari significati di un termine in forma di grafo i cui rami sono tipizzati in base al tipo di relazione.

> Inserendo *"Information"* ad esempio, si può navigare il grafo fino ad arrivare ai termini *"Data"* e *"Knowledge"* da un lato, e *"Entropy"* dall'altro. E ancora, *"Love"* presenta numerosi nodi, che coprono tutti i vari significati del termine, dai significati più astratti all'attrazione fisica, e presenta anche, con un rapporto di antinomia, la parola *"Death"*[67].

L'uso di WordNet con *ThinkMap* dimostra come sia possibile creare servizi avanzati sulla base di contenuti organizzati semanticamente.

4) Intelligenza collettiva

È uno dei tratti essenziali della metodologia e della pratica di ricerca che caratterizza lo sviluppo delle reti telematiche sin dal principio, un fenomeno che Pierre Lévy definisce come "intelligenza collettiva", intelligenza distribuita e connessa in rete che valorizza le competenze individuali attraverso uno spirito collaborativo[68].

Un esempio illustre di intelligenza collettiva è Wikipedia, un'enciclopedia online basata sull'idea che ciascuna voce possa essere aggiunta e modificata da chiunque e che applica alla creazione di contenuti il detto di Eric Raymond (coniato originariamente nel contesto del software *open source*), secondo cui «con molti occhi puntati addosso, ogni bug diventa una bazzecola»[69]. L'impostazione di Wikipedia si sta diffondendo

[63] T. Vander Wal, *La folksonomia e il Web semantico*, in V. Di Bari (a cura di), *Web 2.0*, cit., pp. 78-79.

[64] <http://www.visualthesaurus.com/>.

[65] <http://www.thinkmap.com/>.

[66] <http://wordnet.princeton.edu/>.

[67] F. Meschini, *eContent: tradizionale, semantico o 2.0?*, cit.

[68] P. Lévy, *Intelligenza collettiva. Per un'antropologia del cyberspazio*, Feltrinelli, Milano 1999.

[69] *Ivi*.

anche in ambito scientifico e letterario, dove il tradizionale sistema delle recensioni su carta è affiancato da veri e propri servizi online come quello di Internet Bookshop, Amazon.com (oltre al già citato Anobii, che indicizza oltre undici milioni di libri e, in aggiunta alla possibilità di pubblicare recensioni consente agli utenti di inserire moltissime informazioni come la data di inizio e di fine lettura, commenti come nelle note a margine, un voto che va da uno a quattro)[70].

5) Decentralizzazione: ogni client è anche un server (P2P)

Molti sistemi del web2.0 sono pensati e progettati per incoraggiare la partecipazione diretta degli utenti in almeno tre modi, indicati da Dan Bricklin come segue:

> Il primo, dimostrato da Yahoo!, è di pagare le persone perché lo facciano. Il secondo, che prende ispirazione dalla comunità *open source*, è di cercare volontari che realizzino lo stesso compito. L'Open Directory Project, un concorrente *open source* di Yahoo, ne è il risultato. Ma Napster ha dimostrato un terzo modo. Avendo come default la possibilità di mettere a disposizione automaticamente qualsiasi pezzo musicale che viene scaricato, Napster ha consentito a ogni utente di contribuire all'aumento del valore del database condiviso. Questo stesso approccio è stato seguito da tutti gli altri servizi di condivisione di file P2P.[71]

I protocolli *peer to peer* trasformano i *client* (componenti che accedono ai servizi o alle risorse) in *server* (componenti che forniscono servizi). In pratica, sezioni particolari della memoria dei calcolatori degli utenti divengono accessibili agli altri utenti.

6) Accesso aperto

In tale contesto, le licenze software e il controllo delle API divengono irrilevanti in quanto il software non ha più bisogno di essere distribuito ma solo utilizzato, e il suo valore è proporzionale alla scala e al dinamismo dei dati che esso aiuta a gestire. Il principio dell'accesso aperto permea così tutti gli strati della rete, dai protocolli ai contenuti, ed è un principio che informa la rete ad ogni livello, definendo tanto gli aspetti tecnici quanto la filosofia e gli aspetti socio-culturali del cosiddetto software sociale.

Si osservi infine che dal 2006 è stata introdotta da Jeffrey Zeldman l'espressione Web 3.0[72]. I significati attribuiti al termine a partire da quella

[70] Si veda L. Lipperini, *La critica fai da te. Dai libri al web, ecco il popolo dei recensori*, «la Repubblica», 2 giugno 2009.

[71] D. Bricklin, *The Cornucopia of the Commons*, cit. in T. O'Reilly, *Che cos'è web 2.0*, cit.

data sono stati vari e diversi, e lo stesso Berners-Lee si è espresso al riguardo con le parole che seguono:

> Le persone si chiedono che cosa sia il Web 3.0. Penso che, forse, quando si sarà ottenuta una sovrapposizione della Grafica Vettoriale Scalabile - oggi tutto appare poco nitido, con pieghe ed increspature - nel Web 2.0 e l'accesso ad un Web semantico integrato attraverso una gran quantità di dati, si potrà ottenere l'accesso ad un'incredibile risorsa di dati che genereranno il Web 3.0[73].

In termini più generici, onnicomprensivi e generalmente condivisi, dalla rete dei computer del cosiddetto Web 1.0, passando attraverso le reti sociali che caratterizzano il Web 2.0, si sta ora giungendo a una rete globale di dati. Il Web 3.0 è la sintesi di questi tre aspetti e nel prossimo capitolo lo intenderemo in questa accezione.

[72] J. Zeldman, *Web3.0*, «A List Apart,» 210, 2006, <http://www.alistapart.com/articles/web3point0/>.

[73] Cfr. <http://www.nytimes.com/2006/05/23/technology/23iht-web.html?_r=1>. Trad. it. dalla voce Web3.0 di Wikipedia: <http://it.wikipedia.org/wiki/Web_3.0> (10/09).

CAPITOLO 4

FARE SCIENZA IN RETE

> La ricerca scientifica è stata completamente trasformata dalla disponibilità della rete; in questo caso si intende ovviamente la grande tela intessuta da Internet e dal web attorno al mondo. Il ricercatore scientifico che pone mano a un nuovo progetto non compie quasi più alcuno degli atti che sino a poco tempo addietro costituivano la norma di questo genere di attività: recarsi in biblioteca, consultare libri, scorrere riviste, aprire decine di classificatori repleti di schede compilate a mano. Nella maggior parte dei casi il ricercatore non entra nemmeno negli archivi del proprio computer. Va direttamente in rete, nella quale sa di trovare in abbondanza i documenti, gli articoli, i dati, i risultati di altri esperimenti, le notizie circa le attività di altri laboratori che lo interessano. Dopo qualche tempo passato a scandagliare accuratamente la rete, il ricercatore si metterà in contatto con altri colleghi sparsi nel mondo, al fine di ottenere da essi suggerimenti e commenti – nei limiti, ovviamente, consentiti dalla competizione internazionale tra i ricercatori. Allorché avrà prodotto qualche primo risultato, lo trasmetterà attraverso la rete a qualche collega di cui si fida; e quando considererà sufficientemente maturi i risultati della ricerca li diffonderà in qualche minuto, attraverso la rete, a tutte le riviste specializzate, i media, i laboratori alleati e concorrenti.
>
> L. Gallino, *Tecnologia e democrazia.*

1. Un ritorno all'antico: dalle RFC al software libero

«Un elemento chiave della crescita di Internet è stato l'accesso libero e aperto alla documentazione di base, e in special modo alle specifiche dei protocolli». Ad affermarlo sono i suoi stessi inventori nella già citata *A Brief History of the Internet*, in cui un paragrafo è dedicato in particolare al ruolo della documentazione nello sviluppo e nella diffusione della rete[1]. I giovani laureati che collaborano alla creazione di ARPANET e dei protocolli TCP/IP sono mossi dai medesimi principi che animano il dibattito sulla repubblica delle lettere sin dal suo nascere, un dibattito di cui adottano metodi e ideali: se infatti il concetto di *humanitas* per gli umanisti è strettamente legato alla comunicazione, intesa come un atto di generosità verso gli altri che non si realizza in mera erudizione, comunicazione e comunicare sono termini che vengono usati per indicare il

[1] B.M. Leiner, V.G. Cerf V.G., D.D. Clark, R.E. Kahn, L. Kleinrock, D.C. Lynch, J. Postel, L.G. Roberts, S. Wolff, *A Brief History of the Internet*, cit., p. 35.

Francesca Di Donato, *La scienza e la rete : l'uso pubblico della ragione nell'età del Web*,
ISBN 978-88-8453-494-1 (print) ISBN 978-88-8453-500-9 (online)

dovere di trasmettere il sapere ai posteri; dunque un atto non di mera liberalità, ma di giustizia. Trasmettere le proprie conoscenze costituisce il punto di arrivo e l'anima stessa del lavoro intellettuale: una condizione indispensabile allo svolgimento dell'attività di ricerca. Oltre all'amore per la ricerca sono la volontà e la capacità di comunicare il sapere a definire il dotto degno di questo nome. Il dibattito scientifico pubblico si configura come una rete in progressiva espansione che, attraverso continue citazioni reciproche (che si esprimono in recensioni, note, riassunti, bibliografie e indici, discussioni, lettere aperte) costruisce un discorso comune.

Gli scienziati americani prendono tuttavia le distanze dagli strumenti consueti della comunicazione scientifica, ritenendo il ciclo tradizionale della pubblicazione accademica troppo formalizzato e lento per lo scambio di idee essenziali a creare reti, e si dotano di nuovi mezzi di comunicazione e pubblicazione formalizzando così nuove procedure per la costruzione di consenso, *in primis* le *Request for Comments* (RFC), documenti ufficiali approvati dall'IETF (*Internet Engineering Task Force*) che propongono, aggiornano o dichiarano superati gli standard di Internet[2].

L'invenzione della posta elettronica nel 1972 è uno strumento di comunicazione essenziale. L'e-mail si configura da principio come una *killer application*, un'applicazione decisiva per la diffusione di Internet ed è, ancora oggi, il servizio più usato della rete. La nascita della posta elettronica è molto importante anche per l'affermazione del concetto di *authorship,* letteralmente "paternità intellettuale", concetto che in un primo momento resta indeterminato a vantaggio di una scrittura collettiva e in cui l'autore assume un'importanza marginale e secondaria. La diffusione di *mailing list* integra lo strumento delle RFC sistematizzando le procedure decisionali all'interno dei gruppi di lavoro[3].

Internet, pensata come strumento per condividere l'informazione e le risorse di calcolo, ha rivoluzionato l'attività della ricerca stessa, trasformando i principi della collaborazione tra scienziati e della libera circolazione del sapere scientifico in una pratica di lavoro, e arricchendola nel tempo di sempre nuovi strumenti.

I suoi principi sono stati formalizzati e rafforzati grazie alla sinergia col movimento per il software libero (*free software*) nelle sue diverse varianti[4]. La storia del software libero, che ha il suo epicentro negli Stati

[2] L'archivio delle RFC, curato da Jon Postel, è online all'URL <http://www.ietf.org/rfc.html>. Il progetto di traduzione italiana delle RFC è accessibile al seguente indirizzo <http://rfc.altervista.org/>.

[3] Si osservi ad esempio che la *Internet Engeneering Task Force* ha 75 gruppi di lavoro, ciascuno dei quali lavora su un aspetto particolare della progettazione e implementazione di Internet. I gruppi fanno uso di mailing list per discutere e poi, una volta che sia stato raggiunto consenso su un documento, questo viene distribuito come RFC.

[4] Un'espressione alternativa a *Free software* è *Open source software*, con la quale si intende software il cui codice sorgente è aperto, ma rilasciato con licenze meno re-

Uniti, e in particolare nei laboratori del MIT e in California (che sono state la culla, come abbiamo visto, anche di Internet) ci indica un modello fortunato di sviluppo delle tecnologie informatiche.

È Richard Stallman, sistemista al laboratorio di intelligenza artificiale del MIT dal 1971 al 1983, a dare voce al movimento e a farsi portatore della filosofia del software libero[5] a partire dai primi anni 80.

> La sua convinzione sulla non utilità e, anzi, sulla dannosità di non diffondere il codice di controllo della macchina, basata su premesse insieme etiche e funzionali, trovava una continua conferma nei molti problemi quotidiani connessi all'utilizzo di computer e altra strumentazione elettronica[6].

Così, trovandosi alle prese con una stampante donata dalla Xerox al prestigioso laboratorio di Boston, Stallmann si scontra con l'impossibilità di modificare il codice sorgente del dispositivo a causa dell'imposizione del *copyright* sul software da parte della casa di produzione. Un evento scatenante che porterà il fisico a lasciare il MIT, prefiggendosi l'obbiettivo di difendere la libertà del software e di creare, a tal fine, un sistema operativo aperto e compatibile con Unix, intenzione che comunica al mondo nel 1983 attraverso l'antenata di Internet, ARPANET[7].

Ma che cos'è il software libero? Stallmann ha cura di precisare che l'aggettivo *free* è da intendersi nell'accezione di *free speech* e non di *free*

strittive dalla licenza GNU/General Public License (GPL), promossa dalla Free Software Foundation. Più precisamente, mentre con *free software* si intende una filosofia politica di sviluppo e disseminazione della conoscenza, di cui il software è un caso particolare, l'espressione *open source* corrisponde all'applicazione commerciale dei programmi ispirati a quei principi. Perciò, è comprensibile la diffidenza che i sostenitori del software libero nutrono nei confronti del movimento *open source*. Si veda la seguente definizione di Open Source Software: <http://it.wikipedia.org/wiki/Open_source> (10/09). Cfr. anche C. Di Bona, S. Ockman, M. Stone (a cura di), *Open sources. Voci della rivoluzione Open Source*, Apogeo, Milano 1999 (ed. orig. 1999).

[5] Si veda <http://www.gnu.org/philosophy/philosophy.html>.

[6] M. Berra, A.R. Meo, *Informatica solidale*, cit., p. 83. Per una ricostruzione della storia del movimento GNU – free software, si vedano in particolare le pp. 82-97. Nel 1985 Stallman fonda la Free Software Foundation, un'organizzazione nata allo scopo di convogliare le forze (lavoro e denaro) per lo sviluppo e la diffusione di programmi aperti.

[7] Il sistema operativo è «un insieme di programmi che consentono la gestione razionale delle risorse del calcolatore, dalle unità di ingresso-uscita (tastiera, video, stampante) alla memoria centrale (quella che oggi viene chiamata RAM) e periferica (tipicamente, l'hard disk). Il sistema operativo è la parte centrale del software di base, ossia di quell'insieme di quei moduli software che sono generalmente venduti insieme all'hardware [...]» M. Berra, A.R. Meo, *Informatica solidale*, Bollati Boringhieri Torino 2001, p. 84. Per una definizione tecnica, si veda la voce <http://it.wikipedia.org/wiki/Sistema_operativo>. Unix è un sistema operativo sviluppato dalla AT&T, cfr.<http://it.wikipedia.org/wiki//Unix>.

beer: libero dunque non è sinonimo di gratuito[8]. La definizione di software libero si fonda sull'importante distinzione tra codice sorgente, il programma che viene scritto dai programmatori in linguaggi formalizzati simili al linguaggio naturale, e codice eseguibile, scritto in linguaggio macchina (fatto di zero e uno), molto difficile da scrivere e pressoché impossibile da leggere anche per programmatori molto esperti[9]. Il codice sorgente viene convertito (in termini tecnici "compilato") in codice eseguibile tramite appositi programmi detti appunto compilatori. Un passaggio che consente ai programmi di funzionare, ma che comporta la perdita di informazioni essenziali al fine di comprenderli e modificarli. A sua volta, l'espressione "software libero" include la libertà di eseguire, copiare, distribuire, studiare, cambiare e migliorare il software, azioni possibili solo se, assieme al codice eseguibile di un programma, viene fornito anche il codice sorgente. Pertanto nella definizione di Stallman il software libero è tale se rispetta quattro libertà:

1. Libertà di eseguire il programma, per qualsiasi scopo (libertà 0).
2. Libertà di studiare come funziona il programma e adattarlo alle proprie necessità (libertà 1). L'accesso al codice sorgente ne è un prerequisito.
3. Libertà di ridistribuire copie in modo da aiutare il prossimo (libertà 2).
4. Libertà di migliorare il programma e distribuirne pubblicamente i miglioramenti (e le versioni modificate in genere), in modo tale che tutta la comunità ne tragga beneficio (libertà 3). L'accesso al codice sorgente ne è un prerequisito[10].

Come si desume dalla definizione di cui sopra, nulla vieta che il software libero possa essere venduto.

Nel 1985 Stallman fonda la Free Software Foundation (FSF), un'organizzazione *no profit* basata sul lavoro di volontari[11], per dare vita al progetto GNU, acronimo ricorsivo dell'espressione "GNU is Not Unix" con la quale il programmatore americano intende affermare che il suo sistema operativo sarà compatibile con Unix e tuttavia, a differenza di que-

[8] <http://www.gnu.org/philosophy/free-sw.html>.

[9] «I calcolatori della prima generazione dovevano essere programmati in linguaggio macchina, ossia in questo complicato linguaggio fatto di lunghissime sequenze di simboli 1 e 0. [...] La programmazione in linguaggio macchina era operazione molto lunga e complessa, perché richiedeva la conoscenza minuta dell'architettura del calcolatore, dei codici di tutte le istruzioni e il controllo mentale delle molte celle della memoria centrale. Per questo furono presto ideati linguaggi simbolici e sviluppati opportuni programmi "traduttori", capaci di tradurre il codice simbolico scritto dal programmatore in un programma in linguaggio macchina che facesse le stesse cose». M. Berra, A.R. Meo, *Informatica solidale*, cit., p. 74.

[10] <http://www.gnu.org/philosophy/free-sw.it.html>.

[11] <http://www.fsf.org/>.

st'ultimo, non sarà proprietario e sarà libero dalle restrizioni imposte dal *copyright*. Stallman sceglie inoltre di non rendere il codice sorgente che scrive di pubblico dominio, condizione giuridica in cui rientrano quei beni che non sono suscettibili di appropriazione da parte di alcun soggetto, sia esso pubblico o privato[12], ma di creare un'apposita licenza a garanzia delle quattro libertà enunciate nella definizione di cui sopra e conia a questo scopo il termine *copyleft*. Il termine, che letteralmente significa "permesso di copia", si basa su un gioco di parole che sfrutta la contrapposizione *right/left,* destra e sinistra, e che si traduce praticamente nell'adozione di licenze libere, vale a dire licenze tramite le quali l'autore cede parte dei suoi diritti (economici) sull'opera indicando «ai fruitori dell'opera che essa può essere utilizzata, diffusa e spesso anche modificata liberamente, pur nel rispetto di alcune condizioni essenziali»[13]. La licenza che viene scritta e adottata dalla FSF è la GNU/GPL (General Public License)[14]. La GPL è un documento legale che viene associato al programma e che concede ai licenziatari il permesso di modificarlo, copiarlo e ridistribuirlo con o senza modifiche, gratuitamente o a pagamento. Una sua caratteristica peculiare, che la contraddistingue da altre licenze *copyleft*, è la cosiddetta "viralità":

> Rispetto alle altre licenze di software libero, la GPL è classificabile come "persistente" e "propagativa".
> È "persistente" perché impone un vincolo alla redistribuzione: se l'utente distribuisce copie del software, deve farlo secondo i termini della GPL stessa. In pratica, deve distribuire il testo della GPL assieme al software e corredarlo del codice sorgente o di istruzioni per poterlo ottenere ad un costo nominale. Questa è la caratteristica principe della GPL, il concetto ideato da Richard Stallman e da lui battezzato *copyleft*. Il suo scopo è di mantenere libero un programma una volta che esso sia stato posto sotto GPL, anche se viene migliorato correggendolo e ampliandolo.
> È "propagativa" perché definisce nel testo una particolare interpretazione di "codice derivato", tale che in generale l'unione di un programma coperto da GPL con un altro programma coperto da altra licenza può essere distribuita sotto GPL, o in alternativa non essere distribuita affatto. Nel primo caso si dice che l'altra licenza è "compatibile con la GPL"; nel secondo caso, che non lo è. Questa caratteristica è indicata come *strong copyleft* nella terminologia della FSF. Il suo scopo è evitare che la persistenza venga via via indebolita apportando

[12] Cfr. la relativa voce di wikipedia: <http://it.wikipedia.org/wiki/Pubblico_dominio> (10/09).

[13] <http://it.wikipedia.org/wiki/Copyleft> (10/09).

[14] <http://www.gnu.org/licenses/gpl.html>.

modifiche coperte da un'altra licenza meno libera, inficiando così lo scopo di mantenere libero il software coperto dalla GPL[15].

Dall'ambito ristretto del software, il concetto di *copyleft* invade il campo dei contenuti (testi, brani musicali, video) grazie a Lawrence Lessig, professore di diritto pubblico prima a Stanford e oggi alla Harvard Law School, che nel 2001 fonda il progetto Creative Commons[16].
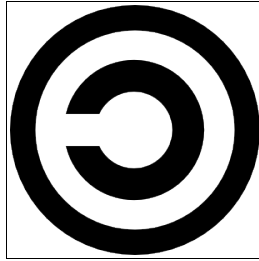
Figura 1. Il simbolo del copyleft, una C rovesciata.

Le licenze Creative Commons permettono di trasmettere parte dei diritti d'autore al pubblico. Nate negli Stati Uniti sono poi state adattate a numerosi sistemi giuridici, compreso quello italiano, dove il diritto d'autore è regolato dalla legge 633/41. Secondo l'art. 6 della legge sul diritto d'autore «il titolo originario dell'acquisto del diritto di autore è costituito dalla creazione dell'opera, quale particolare espressione del lavoro intellettuale» e tutti i diritti gli sono riservati. Viceversa, le licenze stabiliscono in modo esplicito quali diritti sono riservati e quali vengono ceduti.

Le licenze Creative Commons (CC) sono definite dalla combinazione di quattro attributi:

1. Attribuzione: è l'unico attributo obbligatorio, che impone di indicare l'autore dell'opera;
2. Uso commerciale: si autorizzano o meno gli usi commerciali di un'opera;
3. Opere derivate: si autorizza o meno la creazione di opere derivate dall'opera originale (come ad esempio le traduzioni);

15 <http://it.wikipedia.org/wiki/GNU_General_Public_License> (10/09).

16 Cfr. <http://creativecommons.org/>, anche in italiano all'URL <http://www.-creativecommons.it/>. Per le premesse teoriche del progetto di Lessig, si vedano: L. Lessig, *Code and other laws of cyberspace*, Basic books, London 1999; L. Lessig, *Il futuro delle idee,* Feltrinelli, Milano 2006 (ed. orig. 2003); L. Lessig, *Cultura libera, Un equilibrio fra anarchia e controllo, contro l'estremismo della proprietà intellettuale*, Apogeo, 2005 (ed. orig. 2004).

4. Condividi allo stesso modo: qualora si decida di modificare un'opera che ha questa condizione nella licenza, bisogna rilasciarla con la medesima licenza dell'opera originaria.

La combinazione di questi attributi dà luogo alle licenze che seguono:

1. Attribuzione: è necessario citare la fonte originale, mentre è possibile produrre opere derivate senza chiedere il consenso al titolare dei diritti d'autore sull'opera originale e usare l'opera a fini commerciali.

2. Attribuzione-Condividi allo stesso modo: oltre all'obbligo di citare la fonte originale, richiede di applicare la medesima licenza. È consentito l'uso dell'opera a fini commerciali.

3. Attribuzione-Non commerciale: oltre all'obbligo di citare la fonte originale, non consente di usare il testo a fini commerciali.

4. Attribuzione-Non commerciale-Condividi allo stesso modo: oltre all'obbligo di citare la fonte originale e a impedire l'uso dell'opera a fini commerciali, richiede di applicare all'opera derivata la medesima licenza apposta all'opera originale.

Le licenze CC hanno avuto ampia diffusione soprattutto in rete, e non mancano le applicazioni anche in ambito accademico. Il progetto Science Commons, coordinato da John Wilbanks, è finalizzato a questo specifico scopo[17]. L'uso delle licenze Creative Commons favorisce la disseminazione e il riuso della conoscenza.

Si osservi che nel contesto in cui il software libero è nato e si è sviluppato,

> l'etica rappresenta il collante e stabilisce lo standard di un comportamento accettato da chi sente di far parte di questo gruppo in cui la capacità, la conoscenza, l'interesse per il calcolatore sono gli elementi primari della comunicazione e la ragione dello stare insieme, e sono alla base di una organizzazione non gerarchica ma reticolare, basata sullo scambio e sull'interazione[18].

Si diffonde cioè quella che viene definita etica *hacker*, un atteggiamento di cui il mondo accademico può essere considerato il predecessore più antico[19]. *Hacker* è un termine tecnico, la cui definizione è stata spesso travisata e confusa col sabotatore di sistemi informatici (che si definisce, propriamente, *cracker*). Il *Jargon file*, un vocabolario del gergo usato dagli *hacker* e dai professionisti dell'IT, attualmente mantenuto da Eric S. Raymond, contiene le definizioni del termine *hacker* e le regole della *netiquette,* la buona educazione da rispettare in rete[20]. Gli *hacker*

17 <http://sciencecommons.org/>.

18 M. Berra, A.R. Meo, *Informatica solidale*, cit., p. 94.

19 P. Himanen, *L'etica hacker e lo spirito dell'età dell'informazione*, Feltrinelli, Milano 2001, p. 17. Cfr. anche S. Levy, *Hackers. Gli eroi della rivoluzione informatica*, Edizioni Shake Underground, Milano 1996-2002 (ed. orig. 1984).

20 Cfr. <http://www.catb.org/~esr/jargon/>, e, in traduzione italiana, <http://jh-

sono programmatori convinti del fatto che «tutte le informazioni dovrebbero essere libere» e che «l'accesso ai computer [...] dovrebbe essere illimitato e totale»[21], e sono orientati da un'etica del lavoro che affonda le radici nell'accademia e nell'etica del lavoro protestante. Scrive Himanen, citando Weber:

> "Quell'idea peculiare del *dovere professionale*, che oggi è così corrente eppure è tanto poco ovvia, in verità – l'idea di un dovere che l'individuo deve sentire e sente nei confronti del contenuto della sua attività 'professionale', quale che possa essere, e, in particolare, indipendentemente dalla necessità che essa appaia, a una sensibilità ingenua, come pura valorizzazione della propria forza-lavoro o persino solo del suo possesso materiale (come 'capitale')". Weber prosegue: "Indispensabile non è solo un forte senso di responsabilità, lo è, in genere, una mentalità che, almeno durante il lavoro, si liberi dalla continua questione: come, con la massima comodità e la prestazione minima, si possa tuttavia ottenere il salario abituale; e che, invece, svolga il lavoro come se fosse assolutamente fine a se stesso – 'vocazione'"[22].

Esistono, tuttavia, anche importanti differenze tra il modello del vecchio capitalismo e il modello aperto in cui gli *hacker* si riconoscono e operano, distribuendo liberamente i prodotti del loro lavoro affinché siano usati, migliorati e distribuiti da altri. Tale modello si rivela molto efficace anche in pratica, come dimostra il caso Linux, il sistema operativo creato dallo studente finlandese Linus Torvalds nei primi anni novanta. È vero che da principio Torvalds sviluppa il proprio sistema operativo mosso da necessità e curiosità personali, quasi per gioco[23]. Ma dal settembre 1991, quando il giovane *hacker* ne rilascia il codice sorgente in rete, qualcosa cambia. Nell'arco di pochi mesi decine, centinaia, migliaia e, infine, milioni di programmatori contribuiscono allo sviluppo di quello che diventerà il primo sistema operativo libero usando Internet, il Web, gruppi di discussione, email e mailing list per discuterne gli sviluppi, prendere decisioni, aggiungere moduli e migliorare quelli esistenti.

> Il lavoro di sviluppo è anche stato suddiviso in moduli indipendenti, attraverso i quali gruppi di hacker diversi creano versioni in competizione tra loro. Un gruppo composto da Torvalds e altri sviluppatori importanti avrebbe poi deciso quali di queste versioni sarebbero state

anc.altervista.org/jargon/Intro.html> (10/09). Per una buona definizione, si veda la relativa voce di Wikipedia in inglese, <http://en.wikipedia.org/wiki/Jargon_File>.

21 P. Himanen, *L'etica hacker*, cit., p. 5.

22 M. Weber, *L'etica protestante e lo spirito del capitalismo*, Rizzoli, Milano 1991 (ed. orig. 1904-05), p. 77 e pp. 84-85, cit. in P. Himanen, *L'etica hacker*, cit., p. 18.

23 L. Torvalds, D. Diamond, *Rivoluzionario per caso. Come ho creato Linux (solo per divertirmi)*, Garzanti, Milano 2002 (ed. orig. 2001).

incorporate nella versione migliorata di Linux (e, naturalmente, anche la struttura modulare di Linux crebbe gradualmente). Il gruppo di Torvalds non mantiene però alcuna posizione permanente di autorità. Il gruppo conserva la sua autorità soltanto per il tempo in cui le sue scelte corrispondono alle opinioni fatte proprie dalla comunità hacker. Se la scelta fatta dal gruppo-guida si dovesse rivelare meno che aperta, la comunità degli hacker procederebbe a sviluppare il progetto nella propria direzione, bypassando gli ex leader[24].

Si definisce così quello che Raymond definirà il modello bazaar, contrapposto al modello cattedrale[25]. Se il secondo è un sistema di produzione gerarchico in cui un gruppo ristretto di persone dirige e coordina le altre, nel primo, detto appunto modello bazaar, la creazione è aperta a tutti. Si tratta dunque di un modello aperto in quanto tutta l'informazione viene messa a disposizione di chiunque in un processo di condivisione in cui chi partecipa collabora a versioni sempre nuove, sempre migliorate rispetto alle precedenti. La scienza è un *commons*, un bene che può essere goduto da una moltitudine di persone come lo sono le strade o lo erano i prati comuni degli inglesi prima delle recinzioni. Garret Harding in *The Tragedy of Commons*[26], ha sottolineato una caratteristica particolare di tali beni.

Se posso portare le mie bestie al pascolo sul prato comune, rispetto alla mia utilità individuale è per me razionale cercare di sfruttarlo il più possibile, perché è gratis; ma in questo modo esaurisco il pascolo stesso, a danno di tutti gli altri. A chi viaggia piace trovare la strada libera; ma se tutti la usano contemporaneamente, perché è gratis, viaggiare diventa impossibile. Ecco la tragedia dei *commons*: i beni comuni, in quanto vengono usati in comune, tendono a venir sfruttati fino all'esaurimento. Tendono, cioè, se rimangono comuni, a cessare di essere beni[27].

Ma la scienza è un *commons* che sfugge alla sua tragedia, cioè al fatto che, essendo sfruttata da tutti, sia soggetta a esaurirsi.

Una teoria scientifica è un *commons* non competitivo, perché chiunque può apprenderla senza che nessun altro sia depauperato nel suo patrimonio di conoscenze. Una strada o un pascolo sono *commons* competitivi, perché un loro uso incontrollato li deteriora e li impove-

[24] P. Himanen, *L'etica hacker*, cit., pp. 57-58.

[25] E.S. Raymond, *La cattedrale e il bazaar*, Apogeo, Milano 1998 <http://www.apogeonline.com/openpress/cathedral> (edd. origg. 1997-2000).

[26] G. Harding, *The Tragedy of the Commons*, «Science» 162, 1968, pp. 1243-1248, <http://www.sciencemag.org/cgi/content/abstract/162/3859/1243>.

[27] M.C. Pievatolo, *Linux, la GPL e le virtù dimenticate dei beni pubblici,* «Linux Magazine», 2004, <http://bfp.sp.unipi.it/~pievatolo/lm/commons.html>.

> risce. Non bisogna fare l'errore di confondere i pascoli di erba con i pascoli delle idee: le idee, a differenza dell'erba, crescono se vengono condivise, e il loro valore aumenta, perché la condivisione dà loro la possibilità di svilupparsi e di migliorarsi[28].

Gran parte dell'infrastruttura di Internet e del Web, così come i progetti di software *open source* come Linux, è tale che gli utenti, perseguendo propri interessi, costruiscano in modo cumulativo un valore collettivo.

> Ognuno di questi progetti ha un piccolo nucleo centrale, meccanismi di estensione ben definiti e un approccio che consente a chiunque di aggiungere qualsiasi componente ben funzionante, facendo crescere gli strati più esterni di quello che Larry Wall, il creatore di Perl, definisce "la cipolla". In altre parole, queste tecnologie dimostrano gli effetti della rete, semplicemente attraverso il modo in cui sono state progettate[29].

Si osservi che tale modello, pur essendo inclusivo, è anche altamente meritocratico e selettivo. A titolo di esempio vale la pena ricordare il modo in cui la comunità di Debian, una delle più importanti distribuzioni Linux, ha regolamentato la partecipazione al suo sviluppo. Il Contratto Sociale di Debian con la comunità del software libero stabilisce regole rigide che devono seguire coloro che intendono collaborare alla scrittura del software, la prima delle quali ("Debian rimarrà libera al 100%") costituisce il principio fondamentale della comunità. La sua struttura si ispira a quella di una vera e propria società democratica, tanto che il documento che stabilisce la struttura organizzativa dei processi decisionali è chiamato Costituzione. Tutti i partecipanti al progetto sono volontari (senza che ci sia un'azienda di riferimento che lo finanzi) e non esistono né una struttura gerarchica né ruoli dirigenziali, ad eccezione del Debian Project Leader, formalmente a capo del progetto, ma con funzioni principalmente rappresentative e di coordinamento, che è eletto con mandato annuale. Le decisioni più importanti vengono prese a maggioranza dagli sviluppatori Debian, i membri ufficiali del progetto che sono reclutati esclusiva-

[28] *Ibidem*. Si veda, su questo, C. Hess, E. Ostrom, *Ideas, artifacts, and facilities: information as a common-pool resource*, «Law & Contemporary Problems», 66, 2003, pp. 111-146, <http://www.law.duke.edu/shell/cite.pl?66+Law+&+Contemp.+Probs.+111+%28WinterSpring+2003%29>, e in italiano, C. Hess, E. Ostrom (a cura di), *La conoscenza come bene comune. Dalla teoria alla pratica*, Mondadori, Milano 2009 (ed. orig. 2007). Ad ottobre di quest'anno la Ostrom è stata insignita, insieme a Oliver Williamson, del Premio Nobel per l'economia per l'analisi della governance e in particolare dei beni comuni.

[29] T. O'Reilly, *Che cos'è web 2.0*, cit.

mente sulla base di criteri meritocratici (sulla base dei contributi scientifici apportati), ed eletti tramite votazione[30].

Himanen approfondisce infine la differenza tra i due modelli proposti da Raymond sulla base di un'altra analogia, quella tra la Scuola e l'Accademia platonica, dove lo spirito della prima è rappresentato dalla regola monastica di Benedetto[31], mentre la seconda si fonda sulla libertà di critica degli studenti, considerati compagni di apprendimento e non meri fruitori di nozioni. Gli *hacker* sono in ciò molto simili ai filosofi antichi, tanto che

> potremmo anche usare questa visione per creare un'accademia della Rete generalizzata, nella quale tutti i materiali di studio siano liberi di essere usati, criticati e sviluppati da tutti. Migliorando i materiali esistenti in nuove direzioni, il network produrrebbe continuamente migliori risorse per lo studio degli argomenti che di volta in volta si presentano. I membri del network sarebbero guidati dalle loro passioni per i vari temi e dal riconoscimento dei pari per i loro contributi[32].

2. L'avvento del movimento Open Access

Nonostante l'evidente assonanza di questi princìpi con la pratica della comunicazione scientifica tradizionale, fatta eccezione per i casi sopra descritti, il mondo dell'accademia ha tardato a fare propri gli strumenti di comunicazione e, soprattutto, di pubblicazione che Internet e il Web hanno reso disponibili e accessibili. Se infatti è indiscutibile che l'uso dell'email è diventato rapidamente una pratica comune, tra i ricercatori ha prevalso una forma di schizofrenia, che è stata descritta come sindrome del dr. Jeckyill e di Mr. Hyde[33]. Il primo usa attivamente la rete per fare ricerca, sfruttandone mezzi e strumenti quando è a caccia di fonti per scrivere i suoi testi; il secondo, continua a pubblicare con i mezzi tradizionali e ad accettare gli strumenti di valutazione nati e sviluppatisi nel vecchio sistema di comunicazione della scienza, adattandovisi.

È solo tra il 2002 e il 2003, dunque che, sulla scia di alcune significative esperienze pilota[34], in ambito accademico si è cominciato a parlare di

[30] <http://www.debian.org/>. Si vedano anche Il Contratto Sociale Debian <http://www.debian.org/social_contract.it.html> e la Costituzione <http://www.debian.org/devel/constitution.it.html> (in versione italiana).

[31] «Parlare e insegnare spetta al maestro, tacere e ascoltare si addice al discepolo». Benedetto, *La regola di San Benedetto e le regole dei Padri*, Mondadori, Milano 1995.

[32] P. Himanen, *L'etica hacker*, cit., p. 65.

[33] J.-C. Guédon, *Per la pubblicità del sapere,* cit.

[34] Si veda in particolare l'esperienza di ArXiv, l'archivio di pre-print di fisica, matematica, informatica e biologia creato da Paul Ginsparg a Los Alamos nel 1991 e poi

letteratura scientifica O*pen Access*, vale a dire letteratura «digitale, online, gratuita e libera da alcune restrizioni dettate dalle licenze per i diritti di sfruttamento commerciale». Condizioni rese possibili grazie alla rete e al consenso dell'autore (o del titolare dei diritti d'autore)[35].

Il movimento per l'accesso aperto alla letteratura scientifica nasce nella comunità accademica avviando una campagna in favore della condivisione dell'informazione e della conoscenza, intese come beni comuni, e come soluzione al problema che affligge il mondo delle biblioteche e noto come "crisi del prezzo dei periodici", vale a dire l'aumento vertiginoso del costo degli abbonamenti alle riviste scientifiche[36]. Su tale base, l'*Open Access* viene a trovarsi al centro di un ampio dibattito tra ricercatori, bibliotecari, amministratori di università e centri di ricerca, agenzie di finanziamento, editori e studenti.

La fase politica del movimento è preceduta da uno stadio preliminare in cui si definiscono le specifiche tecniche dei software necessari alla creazione di archivi e riviste elettronici. L'archivio di Los Alamos creato da Paul Ginsparg sulla base del riconoscimento del fatto che il bisogno di scambiare risultati scientifici risponde a un fine diverso e complementare al modo in cui funziona il sistema delle pubblicazioni scientifiche, ha fornito lo spunto per la nascita di progetti simili, pur con un duplice sistema organizzativo: agli archivi disciplinari come ArXiv si è accompagnata l'istituzione di archivi istituzionali, vale a dire archivi installati da un'istituzione (tipicamente un'università o un centro di ricerca) per raccogliere i prodotti dei ricercatori che vi afferiscono. «In breve, il movimento ha cominciato a crescere e ad espandersi fino a che è divenuta evidente la necessità di una sorta di azione federativa», necessità che ha trovato risposta a Santa Fe nell'ottobre 1999, quando furono definite le Santa Fe Conventions.

> Lo spirito che presiedette all'incontro ricorda quello che guidava Vint Cerf e Bob Kahn quando progettarono la prima versione di ciò che è diventato il TCP/IP alla fine del 1973: ottimizzare il rapporto tra i risultati e i vincoli tecnici nel tentativo di mantenere i secondi più bassi possibile[37].

L'esito è stato il protocollo *Open Archives Initiative – Protocol for Metadata Harvesting* (OAI-PMH), che rende possibile la raccolta dei metadati (classificati secondo l'ontologia Dublin Core sviluppata dalle biblioteche statunitensi e ampiamente nota in ambito bibliotecario) dei docu-

spostato alla Cornell University. Cfr. <http://it.wikipedia.org/wiki/ArXiv> (10/09) e <http://arxiv.org/>. Oggi ArXiv contiene oltre cinquecentosessantamila articoli.

[35] P. Suber, *Breve introduzione all'accesso aperto*, 2004, <http://www.aepic.it/docs/OA/brief-italian.htm>.

[36] Si veda J.-C. Guédon, *Per la pubblicità del sapere,* cit.

[37] *Ivi*, p. 77.

menti depositati negli archivi Open Access. Nel 2001, la nascita di E-prints a Southampton, un software *open source* per la creazione di archivi digitali, cui sono seguiti il progetto D-Space del MIT, CDS-Ware del Cern, Fedora, e altri ancora, ha permesso il diffondersi di archivi aperti compatibili con OAI-PMH[38]. Lo stesso è avvenuto con software pensati per la creazione e la gestione di riviste digitali e ad accesso aperto (tra cui ricordiamo Open Journal Systems dell'università della British Columbia e l'italiano HyperJournal, creato col contributo di fondi PRIN). Le riviste OA sono oggi oltre quattromila[39].

In pratica, il movimento per l'accesso aperto si dota dunque di due canali:

1. Gli archivi elettronici aperti, che non richiedono che i risultati depositati siano stati preventivamente sottoposti a *peer review* e che possono essere istituzionali (come quelli di università e centri di ricerca), o disciplinari. Tali strumenti rientrano nella cosiddetta strategia verde (*Green Road*) per l'accesso aperto, il cui principale sostenitore è Stevan Harnad, professore di psicologia cognitiva all'Università di Southampton e tra i massimi esponenti del movimento[40]. Tra i più noti archivi disciplinari, oltre al già citato ArXiv, è utile ricordare RePEC, per le scienze economiche e E-LIS, per le scienze biblioteconomiche e dell'informazione. Ad ottobre 2009, la quasi totalità delle università e dei centri di ricerca dei paesi sviluppati si è dotata di archivi istituzionali[41].

2. Le riviste ad accesso aperto, che non sono diverse dalla riviste tradizionali se non in quanto rendono gli articoli da loro selezionati liberamente disponibili al mondo.

> I loro costi consistono nella gestione dei processi editoriali, dal referaggio alla manipolazione dei manoscritti, allo spazio sul server. Le riviste ad accesso aperto coprono i propri costi in un modo molto simile alle radio e alle televisioni: coloro che hanno interesse alla disseminazione dei contenuti pagano anticipatamente i costi di produzione in modo che l'accesso alla lettura sia libero[42].

[38] A ottobre 2009, Open DOAR (Directory of Open Access Repositories) <http://www.opendoar.org/> e il Registry of Open Access Repositories (Roar) <http://roar.eprints.org/> indicizzano oltre 1500 archivi OA.

[39] Si veda la Directory of Open Access Journals (DOAJ) <http://www.doaj.org/>, ospitata presso l'università di Lund.

[40] Gli scritti di Stevan Harnad sono liberamente accessibili online dalla sua homepage: <http://www.ecs.soton.ac.uk/people/harnad>.

[41] G. van Westrienen, C.A. Lynch, *Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005*, «D-Lib Magazine», 11, 9, 2005, <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>. Per una lista aggiornata degli archivi aperti in Europa e nel Nordamerica, si vedano <http://www.opendoar.org/countrylist.php?cContinent=Europe> e <http://www.opendoar.org/countrylist.php?cContinent=North%20America>.

Le riviste rientrano nella strategia definita come strada aurea (*Gold Road*) per l'accesso aperto. Tra queste, vale la pena menzionare PLOSone, uno strumento essenziale per la ricerca biomedica pubblicato dalla Public Library of Science.

Ma quali sono i modelli economici che rendono sostenibile la pubblicazione ad accesso aperto?

> Il modello editoriale tradizionale, nel campo della pubblicazione scientifica, funzionava nel modo seguente:
>
> 1. l'autore cedeva gratuitamente i diritti alla riproduzione e allo sfruttamento economico del proprio testo all'editore;
> 2. l'editore assicurava la disseminazione dell'opera dell'autore favorendone indirettamente la carriera accademica, sulla base del "prestigio" della rivista da lui edita;
> 3. le istituzioni di ricerca pubbliche assicuravano all'editore un ritorno economico o tramite il finanziamento della pubblicazione – in particolare per le monografie – con fondi di ricerca (modello "author pays", molto diffuso in campo umanistico) o tramite abbonamenti da parte di biblioteche (prevalente per le riviste scientifiche).
>
> L'autore accettava questo scambio perché la stampa era una tecnologia costosa, che richiedeva un investimento industriale e personale specializzato[43].

Oggi è possibile disseminare la letteratura scientifica a basso costo usando gli archivi di *preprint*. Ma il problema di come finanziare le riviste, rimane.

> Per quanto riguarda le modalità del loro finanziamento, va ricordato che il modello "author pays", basato sui contributi che il singolo autore versa, individualmente, per ogni pubblicazione, attingendo per lo più a fondi pubblici di ricerca, finisce col penalizzare gli autori di eccellenti contributi scientifici sprovvisti di sufficienti risorse finanziarie – evenienza, questa, tutt'altro che rara in anni di forte contrazione delle risorse destinate alla ricerca[44].

Una soluzione consiste dunque nello spostare i finanziamenti dagli abbonamenti alle pubblicazioni (modello cosiddetto *Institution pays*). Le

[42] P. Suber, *Breve introduzione all'accesso aperto*, 2004, <http://www.aepic.it/docs/OA/brief-italian.htm> (ed. orig. 2004).

[43] CRUI – Commissione Biblioteche, Gruppo OPEN ACCESS, *Riviste ad accesso aperto: linee guida*, 2009, pp. 9-10, <http://www.crui.it/HomePage.aspx?ref=1789>.

[44] *Ivi,* p. 10.

istituzioni accademiche, di fatto, finanziano già le pubblicazioni scientifiche accollandosi i costi degli abbonamenti o coprendo direttamente le spese per la pubblicazione tramite i fondi di ricerca; dunque,

> la pubblicazione ad accesso aperto può diventare economicamente praticabile e addirittura conveniente se le istituzioni accademiche e di ricerca intervengono direttamente, preferibilmente in forma consortile, a sostenere i costi di pubblicazione di riviste di alto profilo scientifico[45].

Si osservino inoltre due fatti: in primo luogo, archivi e riviste non sono alternativi ma complementari; mentre le seconde assolvono al compito della disseminazione e restano strumenti essenziali ai fini della valutazione, i primi consentono la certificazione dei risultati e l'archiviazione di lungo periodo anche di quegli articoli che vengono pubblicati su riviste ad accesso riservato. Il progetto SHERPA/RoMEO[46] indica le politiche in materia di *copyright* dei principali editori commerciali, consentendo agli autori di esercitare al meglio i propri diritti. Lo stesso Elsevier consente l'auto-archiviazione in *repository* OA dopo un periodo di pochi mesi (generalmente sei). In secondo luogo, dev'essere sottolineato il fatto che da tali definizioni restano fuori le monografie, uno strumento di pubblicazione essenziale nelle SSH. Le ragioni di tale assenza derivano probabilmente dal fatto che l'*Open Access* nasce nell'ambito delle STM in risposta alla crisi del prezzo dei periodici che interessa le pubblicazioni su riviste. Non mancano tuttavia iniziative volte a favorire la pubblicazione di monografie digitali ad accesso aperto, spesso accompagnate alla tradizionale stampa su carta, come quelle di Open Humanities Press, e, in Italia, di Firenze University Press[47].

Grazie alla diffusione di archivi e riviste, una sempre maggiore quantità di dati diviene liberamente disponibile sul Web. Ma è la filosofia politica alla base del movimento a essere particolarmente interessante ai fini di questo discorso. Scrivono infatti i promotori della Dichiarazione di Berlino[48].

> La nostra missione di disseminazione della conoscenza è incompleta se l'informazione non è resa largamente e prontamente disponibile alla società. Occorre sostenere nuove possibilità di disseminazione della conoscenza, non solo attraverso le modalità tradizionali ma an-

[45] *Ivi,* p. 11.

[46] <http://www.sherpa.ac.uk/romeo/>.

[47] Si vedano: <http://openhumanitiespress.org/>; <http://www.fupress.com/>.

[48] Nel 2009, la Dichiarazione di Berlino è stata sottoscritta da oltre duecentosessanta istituzioni di ricerca (per un elenco completo dei firmatari, si veda <http://oa.mpg.de/openaccess-berlin/signatories.html>).

> che e sempre più attraverso il paradigma dell'accesso aperto via Internet.
> [...]
> Per mettere in pratica la visione di un'istanza globale ed accessibile del sapere, il Web del futuro dovrà essere sostenibile, interattivo e trasparente. I contenuti ed i mezzi di fruizione (*tools*) dovranno essere compatibili e ad accesso aperto.

Il collegamento tra tali affermazioni, la filosofia del movimento per il software libero e le premesse che hanno portato all'invenzione di Internet e del Web, che i sostenitori dell'accesso aperto estendono ai contenuti scientifici, è evidente. Il Web è refrattario all'uso di sistemi di filtro precedenti alla pubblicazione per costituzione, che è invece una caratteristica dell'accademia come istituzione. Ma oggi una riflessione sui criteri di valutazione è al centro del dibattito di ricercatori, amministratori, finanziatori delle università, e tra questi e i governi.

> La separazione del processo di selezione dai vincoli tecnici ed economici permette di guardare ad esso sotto una nuova luce e di riformularlo in modi che dovrebbero rispondere al bisogno degli scienziati in misura maggiore delle forme di valutazione più vecchie che sono emerse come tacito compromesso tra varie specie di vincoli concorrenti nel mondo della stampa.[49].

Se la tradizione della stampa confonde la disseminazione e la selezione, perché gli alti costi di pubblicazione richiedono di essere selettivi, la pubblicazione elettronica e ad accesso aperto può conciliare questi due aspetti. In altre parole diviene possibile costruire un sistema di comunicazione e di valutazione della scienza aperto e democratico poiché non basato su insiemi chiusi, il cui accesso è regolato da pochi oligopolisti privati. Le comunità di ricerca sono chiamate a reinventarlo.

Per favorire l'affermarsi dell'accesso aperto, che ne è una condizione pregiudiziale, una tappa fondamentale è e sarà la definizione di mandati istituzionali, cioè di regolamenti che obblighino o incentivino i ricercatori che pubblicano i propri risultati con i fondi di una data istituzione pubblica a depositare i prodotti della ricerca nell'archivio della stessa istituzione, o che stabiliscano criteri per incentivare la pubblicazione su riviste *Open Access*. In tal senso, i dati al riguardo lasciano ben sperare per il futuro. Nel 2006, mentre qualche editore comincia a sperimentare politiche OA, vengono sottoscritti oltre 50 *policy* in favore dell'accesso aperto e una decina di mandati. Nell'anno successivo nascono molti progetti di digitalizzazione, pubblicazione, archiviazione e numerosi servizi finanziati con fondi pubblici. Le *policy* sottoscritte dalle principali istituzioni

[49] J.-C. Guédon, *Per la pubblicità del sapere,* cit., p. 81.

pubbliche, associazioni scientifiche e università del mondo in favore dell'accesso aperto ammontano a circa ottanta, mentre i mandati arrivano a venti. Un dato dell'ottobre 2009 ci indica che *policy* e mandati ammontano, ad oggi, a oltre centotrenta. Tra essi, è importante ricordare le linee guida dello European Research Council (ERC) del 2007 e del VII programma quadro dell'Unione Europea, che stabiliscono che tutte le pubblicazioni validate da *peer review* derivanti da progetti di ricerca finanziati con fondi europei dovranno essere rese accessibili in un archivio istituzionale o disciplinare appropriato, entro 6 mesi dalla pubblicazione. Nella stessa direzione si muovono le politiche adottate da CNRS, National Institutes of Health (NIH) negli Stati Uniti, Wellcome Trust, Max Planck Institut, le università di Harvard e di Stanford.

Un passaggio ulteriore e decisivo sono le norme inserite nei sistemi di valutazione della ricerca britannico e australiano, le quali vincolano il processo di valutazione dei prodotti della ricerca al loro deposito in archivi aperti. Collegando gli archivi aperti alle anagrafi della ricerca si stabilisce che il deposito in archivi OA è una precondizione necessaria (o un incentivo) per accedere ai finanziamenti pubblici alla ricerca. Le istituzioni sopra citate sanno benissimo che la sopravvivenza della ricerca che finanziano molto dipende da *come* vengono impiegati. Nonostante la CRUI abbia di recente emanato delle Raccomandazioni in favore dell'accesso aperto, l'Italia resta sotto questo profilo un fanalino di coda[50].
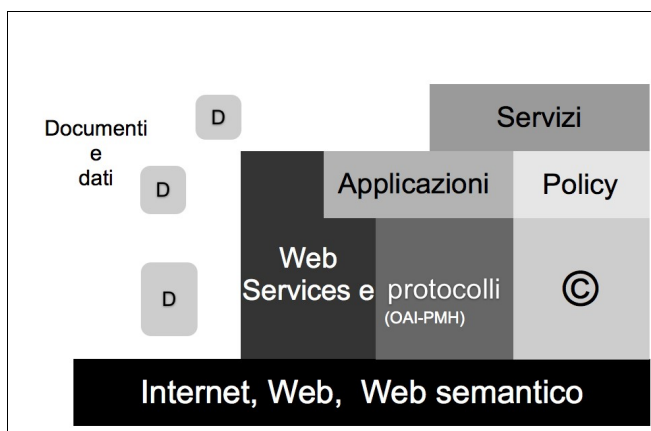
Figura 2. Open Access Layers Cake

50 Per una recente rassegna sull'accesso aperto nel nostro paese, cfr. S. Mornati (a cura di), *Open Access in Italia*, «AIDAinformazioni», 26, 3-4 (fascicolo monografico), 2008, <http://www.aidainformazioni.it/2008/342008monografia.html>. Le Raccomandazioni della CRUI sono disponibili all'URL <http://www.crui.it/HomePage.aspx?ref=1782>.

L'immagine sopra, costruita a mo' di riepilogo sulla scorta della torta a strati del Web semantico, indica gli elementi che costituiscono i tasselli dell'accesso aperto. Alla base si trova l'infrastruttura dell'informazione, Internet e il Web, mentre al piano superiore ci sono due elementi altrettanti fondanti: i protocolli e i servizi web per permettono l'interscambio di metadati tra archivi e riviste ad accesso aperto, e una adeguata gestione dei diritti d'autore. Al terzo piano vediamo le *policy* e le applicazioni grazie a cui è possibile creare archivi e riviste digitali. All'ultimo, invece, si trovano i servizi. Con questo termine intendiamo tutti quegli strumenti che rendono veramente fruibili dati e conoscenze veicolati tramite gli strumenti *Open Access*, al pari di quanto fanno Scopus di Elsevier e Web of Science di ISI-Thomson.

Infatti, è indispensabile sottolineare che all'aumentare dei dati e dei risultati disponibili, sarà sempre più importante l'adozione di sistemi di filtro trasparenti, diffusi e in grado di selezionare l'informazione secondo criteri di qualità molteplici.

In questa direzione si stanno muovendo le piattaforme e i servizi webometrici dei grandi editori e di imponenti progetti di digitalizzazione, primo tra tutti Google Scholar; nella stessa direzione, può andare l'*Open Access*, con il contributo degli strumenti del Web semantico e del Web 2.0.

3. Fare ricerca sul Web 3.0

Ma quali differenze ci sono tra lo spazio a cui dà origine Google Scholar, e la prospettiva offerta da una convergenza tra *Open Access* e Web dei dati?

I critici del colosso di MountainView osservano che alcuni aspetti della politica di Google non possono che presentare problemi. Questi aspetti comprendono il rapporto di Google con la pubblicità, le questioni dell'accentramento, del rispetto della *privacy* degli utenti e della censura dei contenuti.

Relativamente al primo, non si può tralasciare il fatto che Google gestisce e organizza l'informazione traendo da queste operazioni enormi ricavi. Non esiste dunque un problema di conflitto di interessi? Detto in altri termini: come possiamo fidarci dell'oggettività dei risultati delle nostre *query,* se da quei risultati (e dunque dai criteri adottati per selezionare l'informazione) dipendono i profitti di chi fa le ricerche per noi? Che cosa penseremmo se il bibliotecario che prende decisioni sugli acquisti di una biblioteca e sull'organizzazione dei suoi libri, si offrisse, nella medesima biblioteca, di fare pubblicità a uno o più editori, dietro compenso?

Anche la questione dell'accentramento dei servizi messo in atto da Google genera qualche preoccupazione. Una politica tesa ad inglobare applicazioni e servizi sotto un unico marchio aziendale (si pensi alla recente acquisizione di Youtube, che consente a Google di gestire i milioni

di video messi online dagli utenti) presenta il pericolo di una deriva monopolistica che rischia di trasformare l'intelligenza distribuita in un unico grande cervello al cui centro si trova un'azienda privata, mossa da interessi commerciali.

Ma a destare le preoccupazioni maggiori sono i problemi relativi alla tutela della *privacy* e alla censura delle informazioni.

È cosa nota, ammessa dallo stesso Google, che appositi software (detti *crawler*) leggano i nostri documenti online grazie ai loro servizi per fare indagini di mercato ed eseguire la cosiddetta profilazione, vale a dire adattare i risultati delle *nostre* ricerche a quelle che, sulla base del nostro comportamento, il sistema desume che siano le *nostre* preferenze. Il metodo in cui questo avviene è analogo a quello che usa Amazon. Chiunque abbia acquistato un libro su Amazon.com avrà notato che, successivamente, ha cominciato a ricevere email in cui gli si suggeriscono altri libri che potrebbero essere di suo interesse. In pratica, ecco di quali dei nostri dati dispone:

Google
- conosce le nostre ricerche sul web (Google Search);
- conosce la nostra navigazione sul web, i siti che visitiamo più spesso, quanto tempo vi trascorriamo, quali link selezioniamo, e via dicendo (Google Chrome, Google Toolbar);
- conosce le notizie che più ci interessano e i siti di informazione che siamo soliti leggere (Google News);
- sa su quali pubblicità clicchiamo (Google Adwords e Adsense),
- gestisce e organizza milioni di documenti di intranet aziendali (Google Search Appliance e Google Mini);
- conosce i nostri siti (Google Webmaster Tool), sa chi li visita, chi vi acquista e perché (Google Analytics), a volte addirittura li ospita (GooglePages & Blogger);
- conosce le nostre transazioni online e il numero della nostra carta di credito (Google Checkout);
- può leggere la nostra posta elettronica e dunque sa chi sono - almeno in parte - i nostri amici e colleghi (Google Gmail);
- può leggere la "messaggistica istantanea" che scambiamo con amici e colleghi (Google Talk);
- può leggere i nostri documenti (Google Docs & Spreadsheet);
- sa in quali locali vorremmo cenare... il sabato (Google Maps);
- può pure fotografarli (Google Earth, Google Street);
- conosce i nostri interessi culturali e le nostre letture preferite (Google Books Search);
- conosce che cosa ci interessa nel mondo della ricerca accademica (Google Scholar);
- conosce le nostre cartelle cliniche, sa chi è il nostro medico (Google Health);
- sa come vogliamo arredare casa nostra (Google Catalog);
- conosce i nostri appuntamenti (Google Calendar);

- conosce l'intero contenuto dell'hard disk del nostro computer (Google Desktop);
- sa se siamo interessati alla Borsa e a quali azioni (Google Finance);
- sa quali potrebbero essere i nostri prossimi acquisti (Google Froogle);
- sa chi sono i nostri amici, le nostre inclinazioni e i nostri interessi (Orkut);
- sa quali sono le nostre foto preferite (Google Picasa);
- sa quali sono i nostri video preferiti (Google Video, Youtube)[51].

In altre parole, conosce così tante informazioni su di noi che, incrociando anche solo alcuni dei nostri dati, è in grado ottenere un profilo molto preciso. Si tratta di un quadro inquietante, soprattutto alla luce del fatto che è una politica non trasparente, che non lascia all'utente la possibilità di scegliere quali informazioni fare circolare e tra chi.

Anche il problema della censura, il cui esempio più lampante è quello della Cina, ha risvolti preoccupanti. Google Cina ha infatti accettato di censurare pagine "sgradite" al governo locale, una politica che ha giustificato sulla base dell'assunto "meglio poco che niente". Ma siamo sicuri che poco sia meglio di niente? Si osservi che Google rivendica l'oggettività dell'informazione che propone e dei criteri di selezione che adotta. Dichiararsi oggettivi e censurare alcuni siti può essere considerata una politica democratica di accesso all'informazione?

Queste osservazioni ci conducono a riflettere sul modo in cui Google ci permette di ottenere informazioni e, soprattutto, conoscenza. Vincenzo Letta considera tale questione sotto tre profili: oggettività, rilevanza e purezza dei risultati di una *query*.

> Per *oggettività* di un risultato intendo rappresentare il rapporto che esso ha con *una* idea di *verità* socialmente condivisa e se questa possa essere accolta o scambiata in certi casi per *la verità*. Si tratta di un punto estremamente complesso, perché investe concetti fondamentali del nostro pensiero e, sul piano della vitalità e buona salute di una cultura e di una democrazia, oserei dire cruciale. Dovremmo distinguere tra diverse situazioni, prima di tutto fra quelle facilmente oggettivabili e quelle che invece non lo sono. Ad esempio, se voglio sapere chi ha scritto i Promessi Sposi, potrò definire il risultato "Alessandro Manzoni" come vero, ma se voglio trovare il miglior sito che offra informazioni su un periodo storico, ad esempio la Guerra Civile Americana, non è detto che sia vero che i siti che mi vengono segna-

[51] V. Letta, *In Google We Trust*, cit., pp. 38-40. L'elenco dei servizi di Google è aggiornato all'autunno 2008. Di recente (ottobre 2009), Google ha annunciato la nascita di Social Search, un servizio che consente a chi fa una ricerca di restringere i risultati di una *query* alle persone della sua sfera sociale <http://www.google.com/experimental/>.

lati siano i migliori, dove per migliori possiamo intendere sommariamente i più documentati, chiari, liberi e aperti al confronto critico.
Per *rilevanza* intendo quanto il risultato di una ricerca è vicino all'aspettativa di chi ha effettuato quella ricerca. Ad esempio, se effettuo una ricerca con la chiave "apache" con l'intenzione di ricevere informazioni sulla tribù indiana, mentre invece mi vengono segnalati siti che hanno a che fare con il server chiamato Apache, diremmo che quei risultati hanno una bassa rilevanza.
Per *purezza* di un risultato, infine, intendo la sua completa e coerente dipendenza dai criteri pubblici che stanno alla base del funzionamento di un motore, senza che intervengano nascostamente altri fattori. Ad esempio, se io cerco "partiti politici italiani", dovrei ottenere la lista dei siti di tutti i partiti, senza che qualcuno di questi venga escluso perché nascostamente censurato dal motore[52].

In una nota, Letta chiarisce che per criterio pubblico non intende la pubblicità dell'algoritmo (che nel caso di Google, come si è osservato, è segreto) del motore di ricerca, bensì la filosofia che ne ispira il funzionamento.

«Nel caso di Google diremmo che la filosofia di fondo scelta per la valutazione di un sito considera il successo che questo ha ottenuto in termini di link (sia quantitativamente che qualitativamente) come criterio per determinarne la qualità e il peso»[53]. Si tratta di un punto essenziale, poiché è sulla base di tale filosofia che Google valuta la qualità delle risorse e stabilisce un *ranking* tra queste.

Google definisce così come ciò avvenga:

Basato sullo specifico carattere "democratico" del Web, PageRank sfrutta la vastissima rete di collegamenti associati alle singole pagine per determinarne il valore. In pratica, Google interpreta un collegamento dalla pagina A alla pagina B come un "voto" espresso dalla prima in merito alla seconda. Tuttavia, non si limita a calcolare il numero di voti, o collegamenti, assegnati a una pagina. Oltre a effettuare questo calcolo, Google prende in esame la pagina che ha assegnato il voto. I voti espressi da pagine "importanti" hanno più rilevanza e quindi contribuiscono a rendere "importanti" anche le pagine collegate. PageRank assegna ai siti Web importanti e di alta qualità un "voto" più elevato di cui Google tiene conto ogni volta che esegue una ricerca[54].

Dunque,

52 *Ivi*, p. 54.
53 *Ibidem*.
54 <http://www.google.it/intl/it/why_use.html>.

> un link equivale a un voto; tutti i link non hanno lo stesso peso e quelli provenienti da un sito a sua volta molto linkato hanno un peso più alto. In altre parole, la qualità non è nient'altro che un prodotto della quantità, e, spingendosi ancora oltre, è *oggettivo* ciò che è prodotto dall'*opinione* della maggioranza. Si può essere dunque facilmente portati a pensare che in un quadro del genere l'originale, l'atipico, il carattere singolare e tutto ciò che sia decisamente fuori dagli schemi fatichi ad essere intercettato, perché Google privilegia il dato comune, il valore medio, non l'eccezione e ciò che è minoritario. Come scrive Barbara Cassin "L'opinione dunque serve da punto di partenza e di arrivo, è misura e criterio di tutto. Essa definisce lo statuto ontologico degli oggetti che sono sulla ragnatela e della classificazione che ne fa Google"[55].

Si tratta di una definizione di qualità piuttosto discutibile, in quanto rende nullo (perché invisibile o impossibile) il giudizio soggettivo dell'utente, che è colui che, alla fin fine, dovrebbe poter stabilire che cosa è effettivamente valido secondo i *suoi propri* criteri. Confondendo quantità e qualità, e spacciando la prima per la seconda, un progetto come Google Scholar rischia di far sparire interi spazi di conoscenza che, pur periferici, sono di valore. Abbiamo osservato che un motore di ricerca è in grado di rintracciare solo le risorse che si trovano nel nucleo centrale e nel continente out del Web. Le pagine contenute nel primo e nel quarto continente si possono raggiungere solo conoscendone l'URL, e per trovare quanto si trova nel *deep Web* (che è ricco di letteratura scientifica, "grigia" e no) sono necessarie tecniche avanzate di ricerca, che non possono limitarsi all'uso di un motore. Perché dunque dovremmo considerare oggettivo un risultato che non copre l'intero Web e i cui criteri alla base sono oscuri[56]? Inoltre, il modo in cui Google funziona è tale da far sì che ciò che è popolare rimanga tale, mentre ciò che non si conosce, non si conoscerà mai. Ad assumere grande valore nella ricerca scientifica

[55] V. Letta, *In google we trust*, cit., p. 59; B. Cassin, *Google-moi. La deuxième mission de l'Amérique*, Albin Michel, Parigi 2007, p. 104.

[56] «Non può esistere, contrariamente a ciò che hanno sognato gli editori del XIX secolo, la biblioteca universale, ma al massimo dei punti di vista specifici sull'universale. Delle scelte si fanno sempre, obbligatoriamente. Si può valutare in più di cento milioni di opere stampate la produzione totale dell'umanità dopo Gutenberg (mi riferisco solo all'Occidente). La quantità promessa da Google, così impressionante in termini assoluti, corrisponde soltanto a una piccola percentuale di questa immensità. Bisogna quindi chiedersi quali libri verranno scelti e quali criteri interverranno per determinare la selezione. E se Google, consapevole di ciò, annuncia un piano (incerto) di selezione, il dibattito sulla sua validità e il suo eventuale imperialismo è non solo legittimo ma necessario, perché rischia di imporsi a spese dell'eredità di secoli e di saggia riflessione». J.-N. Jeanneney, *Quand Google défie l'Europe. Playdoyer pour un sursaut, Mille et une nuits,* Parigi 2006, p. 11, cit. in V. Letta, *In google we trust*, cit., p. 80.

sono molto spesso testi sconosciuti e introvabili, mentre scoprire qualcosa di nuovo e originale a partire da quello che hanno letto tutti è assai difficile. Infine, se il criterio di seguire l'opinione della maggioranza è adottato da Google, non vale lo stesso per gli altri motori di ricerca; tanto è vero che formulando la stessa *query* su diversi motori, si ottengono risultati molto differenti[57]. Se tutti usassimo solo Google, «l'algoritmo PageRank sarebbe un monarca digitale che governerebbe attraverso una moderna versione della tirannia della maggioranza».[58]

«Il fatto che anche attraverso i nostri click l'algoritmo PageRank determini il valore dei siti [...] è soltanto un sistema di valutazione di cui noi siamo indirettamente partecipi, ma al di fuori di una dimensione collettiva consapevole e strutturata»[59], vale a dire di una comunità scientifica vigile e aperta. Vale infatti la pena ripetere quanto affermava Alessandro Figà-Talamanca nella sua critica all'*Impact Factor*, critica che ben si applica anche a PageRank: «Un giudizio basato su scelte discrezionali oscurate da presunti parametri oggettivi è sottratto ad ogni critica e quindi potenzialmente arbitrario»[60].

In ambito scientifico, questi e altri problemi legati ai progetti di Google rivolti all'accademia, Google Print e Google Scholar, sono stati sottolineati soprattutto dai francesi che, con la collaborazione della Germania, hanno dato vita al programma europeo Europeana finanziato dal VII

[57] Questo sito permette di confrontare i risultati di una stessa *query* ottenuti usando Google e Yahoo <http://www.langreiter.com/exec/yahoo-vs-google.html>.

[58] «[...] sul piano concettuale, è opportuno ricordare i rischi che discendono da questa tirannia della maggioranza, rifacendosi sommariamente al pensiero di colui che introdusse per la prima volta questo concetto nelle sue teorie politiche, ossia Alexis de Tocqueville. Nel confrontare i sistemi politici basati sulla gerarchia tra gli uomini e quello giovane americano, che invece si basava sull'eguaglianza, egli sostiene che non è vero che quest'ultimo si contrapponga ai primi perché in esso gli individui godano di una totale libertà di pensiero. Anche nella democrazia americana esistono costrizioni intellettuali, ma sono più nascoste e hanno una forma diversa da quelle degli stati dispotici. Se nei sistemi dispotici la tirannia si applicava a partire da violenze e costrizioni fisiche, in quelli democratici questa violenza è rimasta, ma è diventata intellettuale. "Non conosco paesi" scrive Tocqueville "dove regni, in generale, meno indipendenza spirituale e reale libertà di discussione che in America (...) In America la maggioranza traccia un cerchio formidabile intorno al pensiero. Dentro questi limiti lo scrittore è libero; ma se osa uscirne cade in disgrazia (...) Il padrone non dice più: penserete come me altrimenti morirete; dice invece: siete liberi di non pensare come me; la vostra vita sarà salva e potrete mantenere i vostri beni; ma da quel giorno sarete stranieri tra noi". Cfr. Alexis de Tocqueville, *Du pouvoir qu'exerce la majorité en Amérique sur la pensée*, in *De la Démocratie en Amérique*, Flammarion, Paris, 1981, tome I, pp. 352-355». V. Letta, *In google we trust*, cit., p. 60.

[59] V. Letta, *In google we trust*, cit., p. 76.

[60] A. Figà-Talamanca, *L'Impact Factor nella valutazione della ricerca e nello sviluppo dell'editoria scientifica*, IV seminario sistema informativo nazionale per la matematica 2000: un modello di sistema informativo nazionale per aree disciplinari, Lecce, 2000, <http://siba2.unile.it/sinm/4sinm/interventi/fig-talam.htm>.

programma quadro. Europeana dovrebbe contenere la digitalizzazione del patrimonio librario, museale e archivistico disseminato in Europa, ma nonostante le buone intenzioni e gli ingenti finanziamenti che ha ricevuto, in pratica stenta ancora a diventare un reale punto di riferimento per i ricercatori, anche per l'assenza di motori di ricerca *ad hoc*[61].

Al contrario, il Web dei *linked data* e le reti sociali che lo abitano possono offrire molto alla comunicazione scientifica, soprattutto se combinati con una politica *Open Access*. Il *Semantic Web* ha bisogno, per funzionare, di dati liberamente accessibili, interpretabili e fruibili. E l'accesso aperto può essere reso effettivamente godibile grazie ai servizi e agli strumenti che la *Web science* sta sviluppando in modo completamente nuovo rispetto a quanto possiamo anche solo immaginare. L'impatto del Web semantico sulla comunicazione scientifica è in grado di trasformare l'attuale paradigma in diverse direzioni, non ultima quella che riconosce la necessità di impostazioni interdisciplinari che siano in grado di incrociare dati e strumenti provenienti da più fonti e discipline[62]: «l'innovazione nelle scienze si genera spesso nelle intersezioni»[63].

> Nei prossimi anni, ci aspettiamo strumenti per la pubblicazione di articoli sul Web che siano in grado di aiutare gli utenti a includere automaticamente negli articoli che producono più *mark*-up comprensibile alle macchine. Mentre gli strumenti correnti che usano XML (*Extensible Markup Language*) possono consentire a un utente di asserire che una parte di un documento è relativo a un 'esperimento', i nuovi linguaggi consentiranno a uno scienziato di esprimere che l'esperimento fa uso di particolari sostanze e reagenti chimici; che il sistema utilizzato riguarda alcune materie organiche particolari; che l'esperimento ha prodotto gel che contengono determinate informazioni sul loro DNA (e che le immagini di questi gel si trovano in luoghi particolari del Web); e così via. […]
> Il Web semantico fornirà le tecnologie sottostanti e unificanti che consentiranno a questi concetti di essere progressivamente collegati in una ragnatela universale di conoscenza (*a universal web of know-*

[61] <http://www.europeana.eu/portal/>. La Francia ha anche cercato di creare un motore di ricerca pubblico, Quero, finanziato nel 2008 dall'EU con quasi 100 milioni di Euro <http://en.wikipedia.org/wiki/Quaero>. Un altro progetto finalizzato alla ricerca di materiali scientifici pubblicati in archivi ad accesso aperto è DRIVER <http://www.driver-repository.eu/>.

[62] J. Hendler, *Reinventing Academic Publishing, Part 1*, «IEEE Intelligent Systems», 2007a, <http://www2.computer.org/portal/web/csdl/doi/10.1109/MIS.2007.93>; J. Hendler, *Reinventing Academic Publishing, Part 2*, «IEEE Intelligent Systems», 2007b, <http://www2.computer.org/portal/web/csdl/abs/html/mags/ex/2007/06/mex2007060002.htm>; J. Hendler, *Reinventing Academic Publishing, Part 3*, «IEEE Intelligent Systems», 2008, <http://www.computer.org/portal/web/csdl/abs/html/mags/ex/2008/01/mex2008010002.htm>.

[63] P. Bourdieu, *Il mestiere di scienziato*, cit., p. 85.

ledge), aiutando così ad abbattere i muri eretti dalla carenza di informazione e consentendo ai ricercatori di trovare e di comprendere i prodotti di altre discipline scientifiche. La nozione di rivista di medicina separata da una rivista di bioinformatica, distinta dagli scritti di fisici, chimici, psicologi e persino dai maestri d'asilo, un giorno diventerà superata al pari di quanto sta accadendo con le riviste cartacee per i nostri studenti[64].

In termini più generali, il processo di selezione può essere reso più democratico se diventa trasparente: innanzitutto, dovrebbe essere valutato tutto quanto viene prodotto, utilizzando magari una scala di valutazione in luogo della mera accettazione o del solo rifiuto.

Nel paradigma tradizionale, sono in pochi a decidere che cosa merita di essere letto: gli autori inviano il documento agli *editor* che lo trasmettono ai valutatori, e i cui giudizi vincolano gli *editor* nel decidere che cosa pubblicare. In un sistema più democratico, reso possibile dalle nuove tecnologie, gli autori possono sottomettere i loro contributi a *editor* e valutatori, ma anche coinvolgere l'intera comunità scientifica in tale attività, fino a sottoporre i testi che producono direttamente al giudizio del pubblico, combinando criteri formali e informali.

Tra i primi rientrano il *peer review* e gli indici quantitativi di valutazione della qualità. Internet e il Web hanno consentito una trasformazione sia del primo che dei secondi.

In primo luogo, al *peer review* tradizionale si sono aggiunte nuove forme di valutazione qualitative.

1. Il *peer review ex post,* una forma di certificazione della qualità applicata a risultati già pubblicati in archivi (tipicamente sui *preprint*, cioè su articoli che non siano stati sottoposti a referaggio) tramite l'uso di *overlay journal* che selezionano il contenuto archiviato nei *repository* (generalmente in forma di *preprint* ovvero di articoli non referati) validandoli.

2. L'*open peer review*, un'altra forma di *peer review* a posteriori applicato a risultati già pubblicati e sottoposti a referaggio. Un esempio consiste nel lasciare la possibilità di commentare il testo ai lettori in modo simile a quanto avviene sui *blog*[65].

64 T. Berners-Lee, J. Hendler, *Scientific publishing on the 'semantic web'*, «Nature debates», 2001, <http://www.nature.com/nature/debates/e-access/Articles/bernerslee.htm>.

65 Cfr. J. Greaves et al, *Nature's Trial of open peer review*; E. Sandewall, *Opening up the process. A hybrid system of peer review*; T. Koop, U. Pöschl, *An open, two-stage peer-review journal*; E. Koonin, L. Landweber, D. Lipman, R. Dignon, *Reviving a culture of scientific debate*; T. Groves, *How can we get the best out of peer review? A recipe for good peer review* (tutti gli articoli citati sono online all'URL del dibattito di «Nature» (2006), <http://www.nature.com/nature/peerreview/debate/index.html>.

3. Il *soft peer review*[66] che consente di usare i metadati collaborativi archiviati su "online reference managers" come CiteUlike e Connotea, con la possibilità di gestire ampie quantità di dati, che sul Web sono già a disposizione. Nuovi strumenti e parametri di valutazione si basano sull'uso di metadati semantici, indicatori di popolarità, sistemi di *rating* assegnati dai lettori (con un procedimento analogo a quello che avviene sul sito Venere.com, che permette di vedere i giudizi che gli utenti hanno dato sugli alberghi disponibili) e annotazioni collaborative da parte degli utenti.

Il *peer review ex post* – che altro non è che un uso pubblico della ragione libero e illimitato – è oggi una reale alternativa al tradizionale *peer review ex ante*; se sommate alle pratiche tradizionali, le nuove forme di valutazione rese possibili dalla rivoluzione mediatica che stiamo vivendo possono essere un impulso in favore di una democratizzazione della comunicazione scientifica - e in particolare di un ampliamento del processo su cui si basa la valutazione dei risultati, che, come è noto, è un momento fondamentale nella costituzione della Repubblica della Scienza.

Riguardo agli indici quantitativi, si diffondono indici che si fondano sull'utilizzo di criteri basati non solo sulle citazioni ma anche su altre metriche, come ad esempio il *download* degli articoli, e che si configurano come strumenti complementari ai primi. In tale direzione si muovono il progetto MESUR (MEtrics from Scholarly Usage of Resources) avviato da Herbert Van de Sompel, Johan Bollen e Marko Rodriguez del Los Alamos National Laboratory[67] e Usage Factors[68], sviluppato dallo United Kingdom Serials Group in collaborazione con COUNTER, entrambi del 2006.

I vantaggi di una bibliometria basata sull'uso sono molteplici:
- l'utilizzo di un articolo può essere tracciato dal momento stesso in cui l'articolo (come preprint o postprint) viene messo in rete, mentre le citazioni sono calcolabili solo dopo un dato lasso di tempo che varia dai tre ai sei mesi successivi alla pubblicazione, ma che si riduce se l'articolo viene depositato in rete prima ancora della sua pubblicazione;
- l'uso può essere misurato per una varietà di documenti, capitoli di monografie, presentazioni, audio, video, dati primari della ricerca e non si limita a misurare l'impatto dell'articolo scientifico;

[66] Cfr. D. Taraborelli, *Soft peer review: social software and distributed scientific evaluation*, 2008, <http://nitens.org/docs/spr_coop08.pdf >.

[67] <http://www.mesur.org/MESUR.html>. J. Bollen, M.A. Rodriguez, H. Van de Sompel, *MESUR, usage-based metrics of scholarly impact*, 2007, <http://www.mesur-.org/Documentation_files/JCDL07_bollen.pdf>.

[68] <http://www.uksg.org/usagefactors >.

- l’uso può riflettere molti differenti aspetti dell’impatto della ricerca scientifica.[69]

Un sistema che consente l'applicazione di queste nuove forme di valutazione è ipotizzato da Brian Whitworth e Rob Friedman e sintetizzato nello schema che segue. Si tratta di un modello intenzionalmente generico, un ibrido tra un archivio elettronico e una rivista elettronica, che consente un doppio canale di pubblicazione (quello tradizionale, basato sulla revisione anonima da parte dei pari e uno diretto, che permetta la pubblicazione immediata) associato a una scala di valutazione di sei gradini che va dall'eccellente allo sconsigliato, e che comprende anche la categoria “not rated” (non valutato)[70].
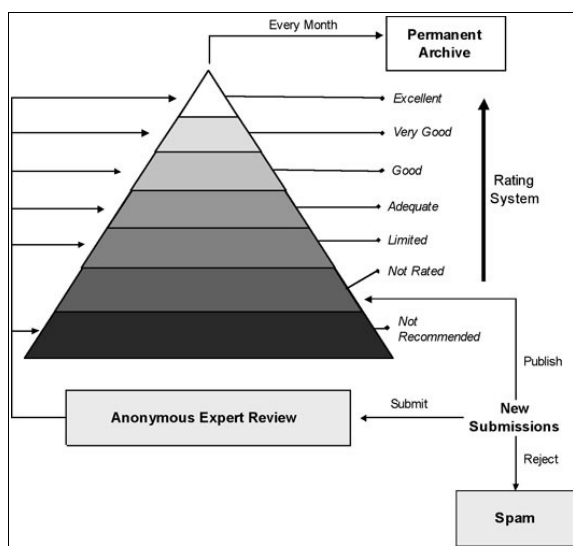
Figura 4. Architettura di un sistema di comunicazione scientifica aperto

Un sistema di filtraggio preliminare potrebbe consistere nell'identificare, tramite la registrazione, i membri dell'accademia in modo che questa qualifica risulti tra le caratteristiche di chi valuta. Ma nel sistema proposto da Whitworth e Friedman i *rating* dei lettori possono sommarsi a quelli degli esperti, lasciando al lettore la scelta tra quali criteri far preva-

69 M. Cassella, *La valutazione della ricerca nelle scienze umane*, cit.

70 B. Whitworth, R. Friedman, *Reinventing Academic Publishing online. PART II. A Socio-Technical Vision*, «First Monday», 14, 9, 2009, <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2642/2287>.

lere e includendo metriche di valutazione sia *ex ante* che *ex post*, in modo analogo a quanto suggerito da Jean-Claude Guédon:

> Invece di essere immediatamente sottoposto a *peer review* e a *copy-editing* per essere pubblicato, l'articolo è semplicemente passato a un vaglio di credibilità e di idoneità. Quindi, viene messo online e offerto all'esame e ai commenti degli specialisti della comunità di riferimento. Le istruzioni e le procedure sviluppate su PLoS One per gli interventi del lettore potrebbero essere facilmente adattate agli archivi, da prima singolarmente e quindi in gruppi. Non sarebbe difficile creare un secondo criterio di valutazione oltre al *peer review* che potrebbe aiutare a valutare il valore di un dato articolo per una determinata comunità. In questo modo, il legame tra i testi contenuti in un archivio e le varie comunità di ricercatori sarebbe molto più forte e navigare tra ricerca all'avanguardia diventerebbe uno sforzo di gruppo molto meglio coordinato[71].

Rendendo accessibile tutto, i criteri di selezione e di valutazione, non più dipendenti dal marchio della rivista o dall'algoritmo della ricerca, possono cambiare in modo che la qualità torni al centro del discorso.

I sistemi di filtraggio resi possibili dal Web semantico sono molto efficaci soprattutto quando l'opinione dell'utente è diversa da quella della media della popolazione. Resta, certo, molto da fare in tal senso. In primo luogo, bisogna identificare le fonti di informazione affidabile, in modo, poi, da sviluppare meccanismi che inferiscano la fiducia in modo accurato da fonti meno sensibili dei grandi connettori. Le relazioni sociali di fiducia, implementate sul Web, possono trasformare il modo in cui gli scienziati usano la rete. Al crescere della quantità di contenuti generati dagli utenti, cresce anche il bisogno di un approccio sociale al filtraggio dell'informazione[72].

A tal fine, l'interoperabilità è una condizione necessaria, ma non è sufficiente. Gli archivi e le riviste digitali non dovranno restare meri strumenti di pubblicazione, ma diventare più usabili, cioè più utili all'attività della ricerca. L'apporto del Web semantico potrà rendere effettivo lo sviluppo del loro potenziale computazionale, una vera rivoluzione nell'attività della ricerca stessa. I servizi che un sistema aperto come quello progettato potrebbe supportare e che consentirebbero di tracciare meglio il valore dei prodotti scientifici sono molteplici. In primo luogo, è essenziale che i materiali sottoposti a *peer review* siano identificati chiaramente come tali, anche grazie a motori di ricerca *ad hoc*. In più, oltre ai sistemi per misurare l'impatto citazionale attraverso indici diversi, e a metriche d'uso, come quelle sviluppate da progetti come MESUR e Usage Factors,

[71] J.-C. Guédon, *Open Access. Contro gli oligopoli nel sapere*, cit., p. 83.

[72] Cfr. J. Golbeck, *Weaving a Web of Trust*, «Science», 321, 5896, 2008, pp. 1640 – 1641, <http://www.sciencemag.org/cgi/content/short/321/5896/1640>.

è importante offrire strumenti che consentano la possibilità di effettuare *peer review ex post*, vale a dire di commentare, correggere, integrare un articolo (possibilità che offre PLOSone) e far visualizzare ai lettori i *rating* di reputazione di autori e valutatori, funzioni del tipo *same again* (come su Amazon), sistemi di bookmark sociali, e strumenti per rendere più efficace il riconoscimento univoco di un autore[73]. Simili criteri di valutazione, basati su forme di accreditamento sociale, possono essere combinati e usati come strumenti per assegnare fondi di ricerca e per decidere in merito al reclutamento e alle promozioni.

Poiché il riconoscimento nell'accademia è fondamentale, quelli proposti possono risultare meccanismi essenziali per indurre i singoli a trarne vantaggio e, assieme, a dare beneficio alla loro comunità, soprattutto in un momento in cui l'università è accusata dall'opinione pubblica di non valorizzare il merito. L'esempio della nascita e dell'evoluzione di Internet, del software libero e del Web ha mostrato e dimostra che è un sistema che funziona; per questo, e non solo perché è giusto, vale la pena esplorarne le possibilità.

73 Alcuni esempi di siti che usano strumenti come quelli descritti sono: <http://pool.newmedia.umaine.edu/index.php>; <http://www.scribd.com/>; <http://pkp.sfu.ca/integration/>; <http://www.myExperiment.org>.

CONCLUSIONI

Alice è una ricercatrice di filosofia e insegna in un'università italiana. Ha pubblicato articoli, saggi e monografie, e ogni anno partecipa a conferenze e convegni in giro per il mondo. Per lavorare, Alice usa la rete.

Lo scorso anno, ad esempio, ha tenuto un corso sul *Conflitto delle facoltà,* per preparare il quale ha tradotto due capitoli dall'Akademie Ausgabe dell'opera di Kant, tutta online e ad accesso aperto, curata e pubblicata dall'università di Bonn[1]. Per lavorare sui testi, ha utilizzato due dizionari tedeschi importanti del tempo: lo Zedler Universal Lexicon e il Deutsche Woerterbuch dei fratelli Grimm, anch'essi completamente online[2]. Ha pubblicato le sue traduzioni su una rivista ad accesso aperto con una licenza Creative Commons (Attribuzione-Condividi allo stesso modo) tale che, se altri studiosi decidono di utilizzarla volendo modificarne alcune parti, possono farlo senza chiederle il permesso (devono solo citarla come fonte originale e pubblicare la loro traduzione alle stesse condizioni). In verità, un professore di un'altra università le ha scritto dicendole che aveva trovato la sua traduzione molto utile; l'aveva utilizzata per un seminario di dottorato su Kant. Voleva solo ringraziarla.

La traduzione della seconda parte del *Conflitto delle facoltà* completa un progetto di ricerca che Alice ha avviato qualche anno prima, e che comprende la traduzione degli scritti politici kantiani. Alice ha pubblicato le traduzioni dei saggi, man mano che erano pronte, su riviste *peer reviewed*, depositandole prima nell'archivio istituzionale della sua università, e anche in appositi archivi disciplinari. Nel tempo, ha potuto verificare quanto sono state utilizzate, controllando il numero di *download* dei suoi testi dai suddetti archivi. Di uno di questi ha pubblicato un'edizione riveduta, tradotta da una collega che l'aveva rilasciata con la medesima licenza. Così, ha risparmiato tempo e lavoro inutile. Perché reiventare la ruota, se c'è già? Si è detta. Per correttezza, ha avvertito la collega, ringraziandola per il lavoro svolto. Non sarebbe stato necessario perché la licenza non lo richiedeva. La collega ha a sua volta risposto: “Sono felice di

[1] <http://www.ikp.uni-bonn.de/kant/> .
[2] <http://www.zedler-lexikon.de/>; <http://germazope.uni-trier.de/Projects/DWB>.

Francesca Di Donato, *La scienza e la rete : l'uso pubblico della ragione nell'età del Web,*
ISBN 978-88-8453-494-1 (print) ISBN 978-88-8453-500-9 (online)

aver contribuito al suo progetto kantiano. [...] Adotterò il suo libro per il mio corso".

Nel frattempo, Alice ha sottoposto il suo progetto al comitato scientifico della collana del suo Dipartimento, che pubblica con Lulu.com. Lulu è un editore online, che offre servizi editoriali su scala globale e a prezzi molto vantaggiosi. Attribuisce gratuitamente l'ISBN ai volumi, e offre servizi di diffusione nei principali canali di distribuzione online (i libri sono inseriti nei cataloghi di Amazon, Barnes&Noble e Internet Book Shop, e nei principali database bibliografici) e attraverso i *social network*. Ma i libri possono essere ordinati anche in libreria. Grazie all'accordo che il suo Dipartimento ha siglato con Lulu, il prezzo di 100 copie del suo libro è sceso a 3 euro a copia, in luogo dei 13 euro che proponeva l'editore tradizionale con cui il Dipartimento pubblicava la sua collana in precedenza.

Per gli studenti, inoltre, Alice ha preparato un commentario agli scritti politici di Kant. È stato un lavoro molto lungo, perché la letteratura secondaria sul tema è estremamente vasta. In questa fase, Alice ha impostato il suo browser nella modalità "lavoro", e ha stabilito alcuni criteri di ricerca *ad hoc*. Alle fonti pubblicate su siti accademici *peer reviewed* ha assegnato il massimo di affidabilità; a un gradino sotto ha collocato gli articoli presenti negli archivi disciplinari e istituzionali, non *peer reviewed*. Inoltre, ha indicato al browser i 4 motori di ricerca su cui impostare le sue *query*, dei quali uno è un motore specifico per le fonti accademiche, e uno è un motore semantico, che cerca i dati nei diversi database aperti sparsi sul Web. Ha impostato il *browser* in modo da ripetere le *query* sui diversi motori, oltre che in italiano, in tedesco, inglese, francese e spagnolo.

Ogni volta che ha trovato qualcosa di utile e interessante l'ha registrato su Citeulike, che ospita la sua bibliografia di lavoro, contrassegnando la singola fonte con parole chiave e indicandola come già letta o da leggere (e con quale urgenza). Quando ha trovato pagine Web non strettamente accademiche, le ha archiviate su Delicious, un sistema di *social bookmarking* che permette di indicizzare le fonti in modo meno strutturato di quanto non faccia Citeulike. Grazie a Citeulike, navigando tramite i *tag* delle folksonomie associati alle sue fonti, ha trovato molti testi, citati da altri ricercatori, che non conosceva, e li ha aggiunti alla sua bibliografia. Alla fine di ogni giornata di lavoro, ha passato in riepilogo i risultati della sua ricerca, ha scremato tra i risultati i meno rilevanti ed ha aggiunto tra le fonti affidabili gli autori i cui articoli ha ritenuto più validi. Così, quando il mattino dopo ha ripreso a lavorare, il browser ha utilizzato questo ulteriore criterio di selezione delle fonti.

Alla fine, Alice ha pubblicato il suo commentario online in forma di ipertesto, inserendo i link alle traduzioni di Kant, e all'originale. Gli studenti hanno così accesso a Kant a più "strati": dal commentario ai testi in traduzione, e dalle traduzioni agli originali in tedesco. Grazie a un wiki, gli studenti possono inoltre inserire commenti e domande, che sono mo-

derati dalla docente e, una volta pubblicati, accessibili a tutti. Un professore di liceo, ad esempio, le ha scritto per dirle di aver usato i suoi materiali didattici per le lezioni.

Una versione cartacea del commentario è stata inserita nella seconda edizione degli scritti politici pubblicata da Lulu. Gli studenti di Alice possono stamparsi direttamente l'ipertesto e le traduzioni dal Web, ma anche comprare il libro ad un prezzo ragionevole, direttamente dal sito di Lulu o attraverso altri canali. Il Sistema bibliotecario dell'Ateneo dell'Università in cui Alice lavora ne ha comprate qualche centinaio di copie a 3 euro, e offre questo servizio agli studenti, vendendole loro al doppio. Così aiuta il Dipartimento a coprire il costo per la gestione della collana.

Alice ha poi scritto un articolo, rivolto a un pubblico più specialistico, illustrando alcune delle scoperte che ha fatto nel corso della sua ricerca e, nell'arco di pochi mesi, è stata invitata a partecipare a un convegno internazionale dedicato a Kant presso l'università di Manchester, dopo che un collega inglese aveva trovato il suo articolo su una rivista *open access* indicizzata su DOAJ.

Lo scorso anno, Alice ha avuto un buon *ranking* nella valutazione della sua ricerca, calcolato sulla base di diversi indicatori che hanno tenuto conto, oltre che della quantità totale delle sue pubblicazioni, del rapporto tra pubblicazioni *peer reviewed* e numero totale, della media dei *download* dei suoi articoli in sei mesi, un anno, due anni e cinque anni, dell'*indice h* calcolato sugli archivi aperti compatibili con OAI-PMH e su Google Scholar, e anche dei giudizi qualitativi che altri ricercatori universitari hanno dato dei suoi testi, aggregati tramite un servizio del CINECA. Una percentuale dei suoi fondi sono destinati a finanziare la collana del Dipartimento su Lulu e a pagare la stampa delle sue pubblicazioni. Ma con quello che le resta potrà fare anche altro: acquistare libri, partecipare alle conferenze che le interessano, finanziare lo sviluppo di software *open source* necessari alla sua attività di ricerca o progetti di digitalizzazione, e molto altro ancora.

Questa storia non è un'utopia filosofica. L'ottanta per cento è una storia vera, perché gli strumenti di cui narra sono già disponibili. Il restante venti per cento è parte della storia di un futuro molto prossimo. Parafrasando la copertina del *Times* del primo gennaio 2007 dedicata agli utenti del Web[3], senza eccessi di ottimismo potremmo dunque concludere dicendo: “Alice sei tu. Sì, proprio tu. Sei tu a controllare la tua ricerca. Benvenuto nel tuo mondo”.

[3] <http://www.time.com/time/covers/0,16641,20061225,00.html>.

INDICE DEI NOMI

Almind, T.C., 30
Amin, M., 27
Arnau Reved, P., 46
Barabási, A.L., 60 e segg., 65
Barbera, M., 12
Baron, S., 16
Benedetto, 101
Berners-Lee, T., 11, 33 e segg., 41 e seg., 51 e segg., 65, 67 e segg., 76 e seg., 79 e seg., 87, 90, 115
Berra, M., 93 e seg., 97
Biagioli, M., 13, 17 e seg.,
Birch, T., 15
Blanchemerle, P., 44
Bollen, J., 116
Bourdieu, P., 22, 114
Boyd Rayward, W., 44, 46
Boyle, R., 14, 17
Bradford, S.C., 23 e seg.
Breccia, A., 12
Bricklin, D., 89
Brin, S., 65
Buckland, M., 46
Bush, V., 43, 46 e segg., 54, 80
Cailliau, R., 42
Carpita, F., 12
Cassella, M., 21, 117
Cassiers, J., 44
Cassin, B., 112
Castells, M., 41
Cerf, V.G., 39, 91
Clark, D.D., 40, 91
Cutter, C.A., 55
D'Alfonso, M., 12
Day, R., 46
Denis, P., 55
De Robbio, A., 23
Dewatripont, M., 21
Dewey, M., 45, 55
Diamond, D., 98
Di Bari, V., 81, 86, 88
Di Bona, C., 93
Dignon, R., 115
D'Iorio, P., 12
Dooley, B., 16
Erdos, P., 61
Erickson, J., 20
Eulero, 59 e seg.
Figà-Talamanca, A., 27 e segg., 113
Flamsteed, J., 14
Friedman, R., 22, 29, 117
Gabler, H.W., 12
Gallino, L., 12, 91
Garfield, E., 24 e segg., 43, 51, 66
Gesner, K., 55
Geurts, P., 20
Ginsparg, P., 101 e seg.
Gnoli, C., 55 e seg., 71
Gohan, G., 55
Golbeck, J., 118
Gore, A., 39
Gori, M., 64
Greaves, J., 115
Greco, P., 47
Groves, T., 115
Gruber, T., 74
Guédon, J.-C., 16, 22, 31, 101 e seg., 106, 118
Hall, W., 11, 68, 70 e seg., 77, 79, 87
Hallet, M., 44
Harding, G., 99
Harnad, S., 103
Héger, P., 43
Heidegger, M., 11
Hendler, J., 11, 68 e segg., 71, 77, 79, 87, 114 e seg.,
Hess, C., 100
van den Heuvel, C., 46
Himanen, P., 97 e segg., 101
Hirsch, J.E., 31

Francesca Di Donato, *La scienza e la rete : l'uso pubblico della ragione nell'età del Web*,
ISBN 978-88-8453-494-1 (print) ISBN 978-88-8453-500-9 (online)

Hobbes, T., 14
Ingwersen, P., 30
Innis, H.A., X, 14
Jacobs, I., 52
Jeanneney, J.-N., 112
Johns, A., 13 e segg., 17
Kahn, R.E., 38 e segg, 91, 102
Kant, I., 11 e seg., 48, 83, 121 e segg.
King, L., 14
Kleinrock, L., 40, 91
Klopstock, F.G., 15
Koonin, E., 115
Koop, T., 115
Kronick, D.A., 13
La Fontaine, H., 43 e segg., 55
Lagoze, C., 20
Landweber, L., 115
Laningham, L., 80
Lassila, O., 69
Leiner, B.M., 40, 91
Lessig, L., 96
Letta, V., 63, 66, 110 e seg.
Lévy, P., 88
Levy, S., 97
Linneo, C., 55
Lipman, D., 115
Lipperini, L., 89
Lynch, C.A., 103
Lynch, D.C., 40, 91
Mabe, M., 27
Manzoni, A., 110
Marini, E., 12
Marino, V., 55, 71
Marx, K., 11
Mattelart, A., 37, 46
Menchelli-Buttini, F., 12
Meo, A.R., 93 e segg., 97
Meschini, F., 85, 88
Metitieri, F., 65
Milgram, S., 60
Mornati, S., 107
Musser, J., 81
Nelson, T., 36, 43, 46, 49
Newton, I., 14
Noruzi, A., 57 e seg.
Numerico, T., 64
Ockman, S., 93
OECD, 21, 23, 30
O'Hara, K., 11, 68, 70 e seg., 77, 79, 87
Oldenburg, H., 15 e segg.
O'Reilly, T., 80, 82 e seg., 84, 89, 100
Ostrom, E., 100
Otlet, P., 43 e segg., 51, 53
Page, L., 65 e seg.
Pareto, V., 23
Pasquinelli, C., 12
Payette, S., 20
Picard, E., 43
Pievatolo, M.C., 12, 19, 99
Pitrelli, N., 47
Platone, X, 48
Pöschl, U., 115
Possenti, I., 12
Postel, J., 40, 91 e seg.
Psaroudakis, I., 12
Quintarelli, E., 86
Ranganathan, S.R., 55 e segg.
Raymond, E.S., 88, 97, 99, 101
Ridi, R., 65
Rieusset-Lemarié, I., 46
Roberts, L.G., 40, 91
Rodriguez, M.A., 116
Roosendaal, H., 20
Rosati, L., 55, 57, 71, 86
Sacchi, S., 21, 23, 27
de Sallo, D., 15
Sandewall, E., 115
Sher, I.H., 24 e seg.,
Socrate, 10
de Solla Price, D., 21, 24, 63
Sprat, T., 14 e seg.
Stone, M., 93
Sturloni, G., 47
Suber, P., 102, 104
Taraborelli, D., 116
de Tocqueville, A., 113
Torpie, R.J., 24 e seg.
Torvalds, L., 98 e seg.
Toscano, M.A., 12
Vander Wal, T., 86 e segg.
Van de Sompel, H., 20, 116
Walsh, N., 52
Warner, S., 20
Weber, M., 98

Weitzner, D.J., XI, 68, 70 e seg., 77, 79, 87
Wesch, M., 70
van Westrienen, G., 103
Whitworth, B., 22, 29, 117
Willinsky, J., 20
Witten, I.H., 64
Wittman, R., 10
Woodmansee, W., 15
Wolff, S., 40, 91
Zeldman, J., 90
Ziman, J., 20, 22, 24
Zurita Sànchez, J.M., 46

BIBLIOGRAFIA

Almind T.C., Ingwersen P., *Infometric analyses on the World Wide Web: Methodological approaches to 'webometrics'*, «Journal of Documentation», 53, 4, 1997, pp. 404–426.

Amin M., Mabe M., Impact Factors: Use and Abuse, «Perspectives in Publishing», 1, 2000, pp. 1-6, <http://www.elsevier.com/homepage/about/ita/editors/perspectives1.pdf>.

Anderson C., *The Long Tail*, «Wired», 12, 10, 2004, <http://www.wired.com/wired/archive/12.10/tail.html>.

Arnau Reved P., *Trascendencia de la vida y obra de Paul Otlet*, <http://www.ucm.es/BUCM/revistas/byd/11321873/articulos/RGID9595220153A.PDF>.

Atkins D., et al, *Revolutionizing Science and Engineering through Cyberinfrastructure*, National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, 2003, <http://www.communitytechnology.org/nsf_ci_report>.

Bailey C.W., *Open Access Bibliography*, 2005-2010, <http://www.digitalscholarship.org/oab/oab.htm>.

Barabási A.L., *Link. La scienza delle reti*, Einaudi, Torino 2004 (ed. orig. 2002).

Barbera M., Di Donato F., *Weaving the Web of Science. HyperJournal and the impact of the Semantic Web on scientific publishing*, «ELPUB2006. Digital Spectrum: Integrating Technology and Culture - Proceedings of the 10th International Conference on Electronic Publishing», 2006, pp. 341-348, <http://elpub.scix.net/cgi-bin/works/Show?204_elpub2006>.

Bates M., *Toward an Integrated Model of Information Seeking and Searching*, Fourth International Conference on Information Needs, Seeking and Use in Different Contexts, 2002, <http://www.gseis.ucla.edu/faculty/bates/articles/info_SeekSearch-i-030329.html>.

Benedetto, *La regola di San Benedetto e le regole dei Padri*, Mondadori, Milano 1995.

Francesca Di Donato, *La scienza e la rete : l'uso pubblico della ragione nell'età del Web*,
ISBN 978-88-8453-494-1 (print) ISBN 978-88-8453-500-9 (online)

Benkler Y., *La ricchezza della rete. La produzione sociale trasforma il mercato e aumenta le libertà*, EGEA Università Bocconi Editore, Milano 2007 (ed. orig. 2006).

Bentivegna S., *Politica e nuove tecnologie della comunicazione,* Laterza, Roma-Bari 2005.

Berners-Lee T., *Information Management: A Proposal*, 1989-1990, <http://www.w3.org/History/1989/proposal.html>.

Berners-Lee T., *L'architettura del nuovo Web*, Feltrinelli, Milano 2001 (ed. orig. 1999).

Berners-Lee T., *Design Issues*, 1993-2009, <http://www.w3.org/DesignIssues>.

Berners-Lee T., Hall W., Hendler J.A., Shadbolt N., Weitzner D., *Creating a Science of the Web*, «Science», 313, 5788, 2006, pp. 769 – 771.

Berners-Lee T., *Original Semantic Web Layers cake*, 2000, <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>.

Berners-Lee T., *The World Wide Web - Past, present, future. Exploring Universality*, 2000, <http://www.w3.org/2002/04/Japan/lecture.html>.

Berners-Lee T., Hendler J.A., Lassila O., *The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities,* «Scientific American Magazine», 2001, <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>.

Berners-Lee T., Hall W., Hendler J.A., O'Hara K., Shadbolt N., Weitzner D.J., *A Framework for Web Science*, Foundations and Trends in Web Science, 1, 1, 2006, <http://eprints.ecs.soton.ac.uk/13347/>.

Berners-Lee T., Hendler J.A., *Scientific publishing on the 'semantic web'*, «Nature», 2001, <http://www.nature.com/nature/debates/e-access/Articles/bernerslee.htm>.

Berra M., *Sociologia delle reti telematiche*, Laterza, Roma-Bari 2008.

Berra M., Meo A.R., *Informatica solidale*, Bollati Boringhieri, Torino 2001.

Berra M., Meo A.R., *Libertà di software, hardware e conoscenza. Informatica solidale 2*, Bollati Boringhieri, Torino 2006.

Biagioli M., *From Book Censorship to Academic Peer Review*, «Emergences», 12, 1, 2002, pp. 91-118.

Birch T., *The History of the Royal Society of London* (1756), reprinted Johnson, New York 1968.

Bollen J., Rodriguez M.A., Van de Sompel H., *MESUR, usage-based metrics of scholarly impact*, 2007, <http://www.mesur.org/Documentation_files/JCDL07_bollen.pdf>.

Borgman C., *Scholarship in the Digital Age. Information, Infrastructure and the Internet*, The MIT Press, London – Cambridge MA 2007.

Bourdieu P., *La spécificité du champ scientifique et les conditions sociales du progrès de la raison*, «Sociologie et sociétés», 7, 1, 1975, pp. 91-118.

Bourdieu P., *Il mestiere di scienziato. Corso al Collège de France 2000-2001*, Feltrinelli, Milano 2003 (ed. orig. 2001).

Boyd Rayward W., *Visions of Xanadu: Paul Otlet (1868-1944) and Hypertext*, «Journal of the American Society of Information Science», 45, 1994, pp. 235-250, <http://people.lis.uiuc.edu/~wrayward/otlet/xanadu.htm>.

Boyd Rayward W., *Paul Otlet Bibliography,* <https://www.ideals.uiuc.edu/html/2142/652/Paul%20Otlet%20Bibiliography.htm>.

Bradford S.C., *Sources of Information on Specific Subjects*, «Engineering: An Illustrated Weekly Journal», 137, 1934.

Bricklin D., *The Cornucopia of the Commons: How to get volunteer labor*, 2001-2006, <http://www.bricklin.com/cornucopia.htm>.

Brin S., Page L., *The anatomy of a Large-Scale Hypertextual Web Search Engine*, <http://infolab.stanford.edu/~backrub/google.html>.

Buchanan M., *L'atomo sociale. Il comportamento umano e le leggi della fisica*, Mondadori, Milano 2008 (ed. orig. 2007).

Buckland M., *What is a Document?*, «Journal of the American Society of Information Science», 48, 9, 1997, pp. 804-809, <http://people.ischool.berkeley.edu/~buckland/whatdoc.html>.

Bush V., *Come possiamo pensare* (ed. orig. 1945), in T. Nelson, *Literary machines 90.1*, Muzzio, Padova 1992, pp. 1/38-53.

Bush V., *Science: The Endless Frontier. A Report to the President*, U.S. Government Printing Office, Washington 1945, <http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm>.

Cassella M., *La valutazione della ricerca nelle scienze umane*, «quaderni CNBA», 11, 2009.

Cassin B., *Google-moi. Le deuxième missione de l'Amerique*, Albin Michel, Paris 2007.

Castells M., *La nascita della società in rete*, Università Bocconi, Milano 2002 (ed. orig. 1996).

Castells M., *Galassia Internet*, Feltrinelli, Milano 2006 (ed. orig. 2001).

Cerf V.G., Kahn R.E., *A protocol for packet network interconnection*, «IEEE Trans. Comm. Tech.», COM 22, V 5, 1974, pp. 627-641.

Connolly D., *A Little History of the World Wide Web*, 2000, <http://www.w3.org/History/>.

Contratto sociale Debian, versione 1.1, 2004,

<www.debian.org/social_contract>.

CRUI – Commissione Biblioteche, Gruppo OPEN ACCESS, *Riviste ad accesso aperto: linee guida*, 2009,<http://www.crui.it/HomePage.aspx?ref=1789>.

CRUI, *OA e la valutazione dei prodotti della ricerca scientifica – Raccomandazioni*, 2009, <http://www.crui.it/HomePage.aspx?ref=1782>.

Day R., *Paul Otlet's book and the writing of social space*, «Journal of the American Society of Information Science», 48, 4, 1997, pp. 310-17.

Denis P., Gohan G., *Storia della biologia*, Einaudi, Torino 1999 (ed. orig. 1997).

De Kerckhove D., Tursi A. (a cura di), *Dopo la democrazia? Il potere e la sfera pubblica nell'età delle reti*, Apogeo, Milano 2006.

De Robbio A., *Analisi citazionale e indicatori bibliometrici nel modello Open Access*, «Bollettino AIB», 47, 3, 2007, <http://eprints.rclis.org/11999/>.

Dewatripont M. et al., *Study on the economic and technical evolution of the scientific publication markets in Europe*, Directorate-General for Research, European Commission, 2006, <http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf>.

Di Bari V. (a cura di), *Web 2.0*, Il Sole 24ore, Milano 2007.

Di Bona C., Ockman S., Stone M. (a cura di), *Open sources. Voci della rivoluzione Open Source*, Apogeo, Milano 1999 (ed. orig. 1999).

Dooley B., Baron S. (a cura di), *The Politics of Information in Early Modern Europe*, Routledge, New York 2001.

EU-TSER project PL97-1296 *The self-organization of the European Information Society* (SOEIS), 1997-99, <http://utopia.duth.gr/~mboudour/soeis/>.

Eulero, *Solutio problematis ad geometriam situs pertinentis*, 1736, <http://math.dartmouth.edu/~euler/docs/originals/E053.pdf>.

Febvre L., Martin H.-J., *La nascita del libro*, Laterza, Roma-Bari 2007 (ed. orig. 1971).

Figà-Talamanca A., *L'Impact Factor nella valutazione della ricerca e nello sviluppo dell'editoria scientifica*, IV seminario sistema informativo nazionale per la matematica 2000: un modello di sistema informativo nazionale per aree disciplinari, Lecce 2000, <http://siba2.unile.it/sinm/4sinm/interventi/fig-talam.htm>.

Gallino L., *Tecnologia e democrazia. Conoscenze tecniche e scientifiche come beni pubblici*, Einaudi, Torino 2007.

Garfield E., *Citation Indexes for Science. A New Dimension in Documentation through Association of Ideas*, «Science», 122, 3159, 1955, pp.

108-11 (ristampato in «Essays of an Information Scientist», 6, 1983, pp. 468-471), <http://www.garfield.library.upenn.edu/essays/v7-p525y1984.pdf>.

Garfield E., Sher I.H., Torpie R.J., *The use of citation data in writing the history of science*, 1964, <http://www.garfield.library.upenn.edu/papers/useofcitdatawritinghistofsci.pdf>.

Gladwell M., *Il punto critico. I grandi effetti dei piccoli cambiamenti*, Rizzoli, Milano 2000 (ed. orig. 2000).

Gnoli C., *Il tavolino di Ranganathan*, «Bibliotime», III, 3, 2000, <http://www2.spbo.unibo.it/bibliotime/num-iii-3/gnoli.htm>.

Gnoli C., Marino V., Rosati L., *Organizzare la conoscenza. Dalle biblioteche all'architettura dell'informazione*, Tecniche Nuove, Milano 2006.

Golbeck J., Weaving a Web of Trust, «Science», 321, 5896, 2008, <http://www.sciencemag.org/cgi/content/short/321/5896/1640>, pp. 1640 – 1641.

Grabner-Kräuter S., Kaluscha E.A., *Empirical research in on-line trust: A review and critical assessment*, «International Journal of Human-Computer Studies», 58, 2003, pp. 783–812.

Greaves J. et al, *Nature's Trial of open peer review,* «Nature», 2006 <http://www.nature.com/nature/peerreview/debate/index.html>.

Greco P., *Il modello Venezia. La comunicazione nell'era post-accademica della scienza*, in N. Pitrelli, G. Sturloni (a cura di), *La comunicazione della scienza, Atti del I e II convegno nazionale*, Zadig, Milano-Roma, 2004, <http://ics.sissa.it/conferences/csIntroduzione.pdf>.

Groves T., *How can we get the best out of peer review? A recipe for good peer review,* «Nature», 2006, <http://www.nature.com/nature/peerreview/debate/index.html>.

Gruber T., *A translation approach to formal ontologies*, «Knowledge Acquisition», 5, 25, 1993, pp. 199–200, <http://ksl-web.stanford.edu/KSL-92-71.html>.

Guédon J.-C., *Per la pubblicità del sapere. I bibliotecari, i ricercatori, gli editori e il controllo dell'editoria scientifica*, PLUS - Méthexis, Pisa 2004, <http://bfp.sp.unipi.it/ebooks/guedon.html> (ed. orig. 2001).

Guédon J.-C., *Open Access. Contro gli oligopoli nel sapere* (ed. orig. 2008), (a cura di F. Di Donato), ETS, Pisa 2009, <http://www.edizioniets.com/Scheda.asp?N=9788846725172> .

Harding G., *The Tragedy of the Commons*, «Science» 162, 1968, pp. 1243-1248,<http://www.sciencemag.org/cgi/content/abstract/162/3859/1243>.

Hendler J.A., *Reinventing Academic Publishing, Part 1*, «IEEE Intelligent Systems», 2007a, <http://www2.computer.org/portal/web/csdl /doi/10.1109/MIS.2007.93>.

Hendler J.A., *Reinventing Academic Publishing, Part 2*, «IEEE Intelligent Systems», 2007b, <http://www2.computer.org/portal/web/csdl/abs/html/mags/ex/2007/06/mex2007060002.htm>.

Hendler J.A., *Reinventing Academic Publishing, Part 3*, «IEEE Intelligent Systems», 2008, <http://www.computer.org/portal/web/csdl/abs/ht-ml/mags/ex/2008/01/mex2008010002.htm>.

Hendler J.A., Shaboldt N., Hall W., Berners-Lee T., Weitzner D., *Web Science: An Interdisciplinary approach to understanding the World Wide Web*, «Communications of the ACM», 51, 7, 2008, pp. 60-69, <http://csi-india.org/web-science-interdisciplinary-approach-under-standing-web>.

Henry G., *On-line publishing in the 21-st Century: Challenges and Opportunities*, «D-Lib Magazine», 9, 10, 2003, <doi:10.1045/october2003-henry>.

Hess C., Ostrom E., *Ideas, artifacts, and facilities: information as a common-pool resource*, «Law & Contemporary Problems», 66, 2003, pp. 111-146, <http://www.law.duke.edu/shell/cite.pl?66+Law+&+Contemp.+Probs.+111+%28WinterSpring+2003%29>.

Hess C., Ostrom E. (a cura di), *La conoscenza come bene comune. Dalla teoria alla pratica*, Mondadori, Milano 2009 (ed. orig. 2007).

van den Heuvel C., *Mundaneum. Architectures of Global Knowledge: The Mundaneum and the World Wide Web*, «Destination Library», 15, 2008, pp. 48-53, <http://www.virtualknowledgestudio.nl/staff/charles-van-den-heuvel/vdheuvel-mundaneum.pdf>.

Himanen P., *L'etica hacker e lo spirito dell'età dell'informazione*, Feltrinelli, Milano 2001 (ed. orig. 2001).

Hirsch J.E., *An index to quantify an individual's scientific research output*, 2005, <http://arxiv.org/abs/physics/0508025>.

Innis H.A., *Le tendenze della comunicazione*, SugarCo, Milano 1982 (ed. orig. 1951).

Ippolita, *Luci e ombre di Google. Futuro e passato dell'industria dei metadati*, Feltrinelli, Milano 2007, <http://ippolita.net/google>.

Jacobs I., Walsh N. (a cura di), *Architecture of the World Wide Web. Volume One*, 2004, <http://www.w3.org/TR/webarch/>.

Jeanneney J.-N., *Quand Google défie l'Europe. Playdoyer pour un sursaut,* Mille et une nuits, Parigi 2006.

Johns A., *The Nature of the Book. Print and Knowledge in the Making,* Chicago University Press, Chicago 1998.

Kahn R., *Communications Principles for Operating Systems,* «Internal BBN memorandum», 1972.

Kant I., *Critica della ragion pura*, Utet, Torino 1967 (ed. orig. 1781-87).

Kant I., *Risposta alla domanda: Che cos'è l'illuminismo?* (tr. it. di F. Di Donato), «Bollettino telematico di filosofia politica», <http://bfp.sp.unipi.it/classici/illu.html> (ed. orig. 1784).

King L. (a cura di), *The Life and Letters of John Locke*, Garland, New York, 1984 (ed. orig. 1864).

Koonin E., Landweber L., Lipman D., Dignon R., *Reviving a culture of scientific debate,* «Nature», 2006, <http://www.nature.com/nature/peerreview/debate/index.html>.

Koop T., Pöschl U., *An open, two-stage peer-review journal,* «Nature», 2006, <http://www.nature.com/nature/peerreview/debate/index.html>.

Kronick D.A., *Peer review in 18th century scientific journalism*, «JAMA», 263, 10, 1990.

Laningham S. intervista Tim Berners-Lee, 22 agosto 2006, <http://www.ibm.com/developerworks/podcast/dwi/cmint082206txt.html>.

Leiner B.M., Cerf V.G., Clark D.D., Kahn R.E., Kleinrock L., Lynch D.C., Postel J., Roberts L.G., Wolf S., *A Brief History of the Internet*, version 3.32, 2003, <http://www.isoc.org/internet/history/brief.shtml>.

Lessig L., *Code and other laws of cyberspace*, Basic books, London 1999.

Lessig L., *Il futuro delle idee,* Feltrinelli, Milano 2006 (ed. orig. 2003).

Lessig L., *Cultura libera, Un equilibrio fra anarchia e controllo, contro l'estremismo della proprietà intellettuale*, Apogeo, Milano 2005 (ed. orig. 2004).

Letta V., *In Google we trust. Storia, vizi e virtù di un motore di ricerca nordamericano*, Tesi di laurea in etica della comunicazione, Facoltà di Lettere e filosofia dell'Università di Pisa, 2008, <http://etd.adm.unipi.it/theses/available/etd-10132008-104502/>.

Lévy P., *Intelligenza collettiva. Per un'antropologia del cyberspazio*, Feltrinelli, Milano 1999 (ed. orig. 1994).

Levy S., *Hackers. Gli eroi della rivoluzione informatica*, Edizioni Shake Underground, Milano 1996-2002 (ed. orig. 1984).

Licklider J.C.R., Clark D.D., *On-Line Man Computer Communication*, 1962.

Lipperini L., *La critica fai da te. Dai libri al web, ecco il popolo dei recensori*, «la Repubblica», 2 giugno 2009.

Marchionini G., *From Information Retrieval to Information Interaction*, 26th Annual European Conference on Information Retrieval, 2004, <http://ils.unc.edu/~march/ECIR.pdf>.

Mattelart A., *Storia della società dell'informazione*, Einaudi, Torino 2002 (ed. orig. 2001).

Merton R.K., *Sulle spalle dei giganti*, Il Mulino, Bologna 1991.

Meschini F., *eContent: tradizionale, semantico o 2.0?*, 2006, <http://dspace.unitus.it/handle/2067/162>.

Metitieri F., Ridi R., *Biblioteche in rete. Istruzioni per l'uso*, Laterza, Roma-Bari 2005.

Milgram S., *The "Small World" Problem*, «Psychology Today», 1, 1967.

Mornati S. (a cura di), *Open Access in Italia*, in «AIDAinformazioni», 26 (2008), 3-4 (fascicolo monografico) <http://www.aidainformazioni.it/2008/342008monografia.html>.

Morville P., *Information Architecture 3.0*, 2006, <http://semanticstudios.-com/publications/semantics/000149.php>.

Musser J., *Web Service e mashup*, in V. Di Bari (a cura di), *Web 2.0*, Il Sole 24 ore, Milano 2007.

Nelson T., *A File Structure for the Complex, The Changing and the Indeterminate*, «ACM 20th National Conference», 1965, pp. 84-100.

Nelson T., *Literary machines 90.1. Il progetto Xanadu*, Muzzio, Padova 1992, (edd. origg. 1981, 1987, 1990, 1993).

Noruzi A., *Application of Ranganathan's Laws to the Web*, «Webology», 1, 2, 2004, <http://www.webology.ir/2004/v1n2/a8.html>.

OECD, *The measurement of scientific and technical activities: R&D statistics and output measurement in the higher education sector*, Paris, 1990.

OECD, *Report on scientific publishing of 2004-2005, DSTI/ICCP/IE(2004) 11/FINAL*, <http://www.oecd.org/dataoecd/42/12/35393145.pdf>.

O'Reilly T., *Che cos'è web 2.0*, 2004 (ed. orig. 2004), <http://www.aware-design.eu/articles/14-Cos-Web-2-0> .

Otlet P., Blanchemerle P., Cassiers J., Hallet M., *Le Sommaire périodique des revues de droit: Table mensuelle de tous les articles et études juridiques publiés dans les périodiques belges et étrangers,* Veuve Ferdinand Larcier, Bruxelles 1891-1893.

Otlet P., *Un Peu de Bibliographie*, Palais, 1891-92.

Otlet P., *Traité de documentation. Le livre sur le livre. Théorie et pratique*, 1934, <https://archive.ugent.be/handle/1854/5612>.

Otlet P., *Monde: Essai d'universalisme*, 1935, <https://archive.ugent.be/handle/1854/8321>.

Pievatolo M.C., *Linux, la GPL e le virtù dimenticate dei beni pubblici,* «Linux Magazine», 11, 2004, <http://bfp.sp.unipi.it/~pievatolo/lm/commons.html>.

Pievatolo M.C., *La comunicazione del sapere. La questione del diritto d'autore*, «Bollettino telematico di filosofia politica», 2007-2008, <http://bfp.sp.unipi.it/dida/fpa/>.

Platone, *Menone,* Laterza, Roma-Bari, 2004.

Platone, *Fedro*, (tr. it. di P. Pucci), Laterza, Roma-Bari 1966-2000.

Quintarelli E., *Folksonomies: power to the people*, «ISKO Italy-UniMIB meeting», 2005, <http://www.iskoi.org/doc/folksonomies.htm>.

Ranganathan S.R., *Five Laws of Library Science*, Madras Library Association Madras, - Edward Goldston, London 1931, <http://dlist.sir.arizona.edu/1220/>.

Ranganathan S.R., *Prolegomena to Library Classification*, Asia Publishing House, New York, 1967, <http://dlist.sir.arizona.edu/1151/>.

Raymond E.S., *La cattedrale e il bazaar*, Apogeo, Milano 1998 (ed. orig. 2004), <http://www.apogeonline.com/openpress/cathedral>.

Rieusset-Lemarié I., *P. Otlet's Mundaneum and the International Perspective in the History of Documentation and Information Science*, «Journal of the American Society of Information Science», 48, 4, 1997, pp. 301-310.

Roberts L., Merrill T., *Toward a Cooperative Network of Time-Shared Computers,* «Fall AFIPS Conf.», 1966.

Roberts L., *Multiple Computer Networks and Intercomputer Communication*, «ACM Gatlinburg Conf.», 1967.

Roosendaal H., Geurts P., *Forces and functions in scientific communication: an analysis of their interplay,* «Cooperative Research Information Systems in Physics», 1997, <http://www.physik.unioldenburg.de/conferences/crisp97/roosendaal.html>.

Rosati L., *La ricerca si fa sociale: il tagging e la folksonomia*, in V. Di Bari (a cura di), *Web 2.0*, Il Sole 24ore, Milano 2007.

Rosati L., *Architetture dell'informazione. Trovabilità: dagli oggetti quotidiani al Web*, Apogeo, Milano 2007.

Sacchi S., *Comunicazione scientifica e open access. Problematiche sociali e tecnologiche nell'applicazione in Italia*, tesi di laurea, 2003-04, <http://dspaceunipr.cilea.it/bitstream/1889/359/2/comunicazione_scientifica_e_open_access_simone_sacchi.pdf>.

Sandewall E., *Opening up the process. A hybrid system of peer review,* «Nature», 2006, <http://www.nature.com/nature/peerreview/debate/index.html>.

Shaboldt N., Berners-Lee T., *Web science emerges*, «Scientific American», 2008, <http://www.sciamdigital.com/index.cfm?fa=Products.ViewIssuePreview&ARTICLEID_CHAR=47CB5956-3048-8A5E-10A6-C89135BB5E2F>.

de Solla Price D., *Networks of Scientific Papers. The Pattern of bibliographic references indicates the nature of the scientific research front*, «Science», 149, 30, 1965, pp. 510-15, <http://www.garfield.library.upenn.edu/papers/pricenetworks1965.pdf>.

Sprat T., *History of the Royal Society*, Martyn, London 1667.

Suber P., *Breve introduzione all'accesso aperto*, 2004 (ed. orig. 2004), <http://www.aepic.it/docs/OA/brief-italian.htm>.

Tanenbaum A.S., *Reti di calcolatori*, Pearson Education Italia, Milano 2003 (ed. orig. 2002).

Taraborelli D., *Soft peer review: social software and distributed scientific evaluation*, 2008, <http://nitens.org/docs/spr_coop08.pdf >.

de Tocqueville A., *De la Démocratie en Amérique*, Flammarion, Paris 1981.

Torvalds L., Diamond D., *Rivoluzionario per caso. Come ho creato Linux (solo per divertirmi)*, Garzanti, Milano 2002 (ed. orig. 2001).

Vander Wal T., *La folksonomia e il Web semantico*, in V. Di Bari (a cura di), *Web 2.0*, Il Sole 24ore, Milano 2007.

Van de Sompel H., Payette S., Erickson J., Lagoze C., Warner S., *Rethinking Scholarly Communication: Building the System that Scholars Deserve*, «D-Lib Magazine», 10, 9, 2004, <http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html>.

Weber M., *La scienza come professione*, Einaudi, Torino 2004 (ed. orig. 1919).

Weber M., *L'etica protestante e lo spirito del capitalismo*, Rizzoli, Milano 1991 (ed. orig. 1904-05).

Weitzner D.J., Hendler J.A., Berners-Lee T., Connolly D., *Creating a Policy-Aware Web: Discretionary, rule-based access for the World Wide Web*, in E. Ferrari B. Thuraisingham (a cura di), *Web and Information Security*, Idea Group Inc, Hershey PA 2005, <http://www.mindswap.org/users/hendler/2004/PAW.html>.

Wesch M., *Web 2.0 ... The Machine is Us/ing Us*, <http://www.youtube.com/watch?v=6gmP4nk0EOE>.

Wesch M., *Antropologia del Learning 2.0,* in V. Di Bari (a cura di), *Web 2.0*, Il Sole 24ore, Milano 2007.

van Westrienen G., Lynch C.A., *Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005*, «D-Lib Magazine», 11, 9, 2005, <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>.

Whitworth B., Friedman R., *Reinventing Academic Publishing online. PART I. Rigor, Relevance and Practice*, «First Monday», 14, 8, 2009, <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2609/2248>.

Whitworth B., Friedman R., *Reinventing Academic Publishing online. PART II. A Socio-Technical Vision*, «First Monday», 14, 9, 2009, <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2642/2287>.

Willinsky J., *The Access Principle. The Case for Open Access to Research and Scholarship*, The MIT Press, London – Cambridge MA 2006.

Willinsky J., *Proposing a Knowledge Exchange Model for Scholarly Publishing*, «Current Issues in Education», 2000.

Witten I.H., Gori M., Numerico T., *Web Dragons. Inside the Myths of Search Engine Technology*, Morgan Kaufmann Publishers-Elsevier, San Francisco 2007.

Wittman R. (a cura di), *Nachdruck und geistiges Eigentum*, Kraus International Publications, München 1981.

Woodmansee M., *The Genius and the Copyright: Economic and Legal Conditions of the Emergence of the 'Author'*, «Eighteenth-Century Studies», 17, 4, 1984, pp. 425-448, <http://www.compilerpress.atfreeweb.com/Anno%20Woodmansee%20Genius%20&%20Copyright.htm>.

Zeldman J., *Web3.0*, «A List Apart,» 210, 2006, <http://www.alistapart.com/articles/web3point0/>.

Ziman J., *Il lavoro dello scienziato: gli aspetti filosofici e sociali della scienza e della tecnologia*, Laterza, Roma-Bari 1987 (ed. orig. 1984).

Zurita Sànchez J.M., *El paradigma otletiano como base de un modelo para la organizaciòn y difusiòn del conocimiento cientìfico*, Tesi di laurea in biblioteconomia, Universidad nacional autònoma de México, 2001, <http://etdindividuals.dlib.vt.edu:9090/358/1/paradigma_otletiano.pdf>.

STRUMENTI
PER LA DIDATTICA E LA RICERCA

1. Brunetto Chiarelli, Renzo Bigazzi, Luca Sineo (a cura di), *Alia: Antropologia di una comunità dell'entroterra siciliano*
2. Vincenzo Cavaliere, Dario Rosini, *Da amministratore a manager. Il dirigente pubblico nella gestione del personale: esperienze a confronto*
3. Carlo Biagini, *Information technology ed automazione del progetto*
4. Cosimo Chiarelli, Walter Pasini (a cura di), *Paolo Mantegazza. Medico, antropologo, viaggiatore*
5. Luca Solari, *Topics in Fluvial and Lagoon Morphodynamics*
6. Salvatore Cesario, Chiara Fredianelli, Alessandro Remorini, *Un pacchetto evidence based di tecniche cognitivo-comportamentali sui generis*
7. Marco Masseti, *Uomini e (non solo) topi. Gli animali domestici e la fauna antropocora*
8. Simone Margherini (a cura di), *BIL Bibliografia Informatizzata Leopardiana 1815-1999: manuale d'uso ver. 1.0*
9. Paolo Puma, *Disegno dell'architettura. Appunti per la didattica*
10. Antonio Calvani (a cura di), *Innovazione tecnologica e cambiamento dell'università. Verso l'università virtuale*
11. Leonardo Casini, Enrico Marone, Silvio Menghini, *La riforma della Politica Agricola Comunitaria e la filiera olivicolo-olearia italiana*
12. Salvatore Cesario, *L'ultima a dover morire è la speranza. Tentativi di narrativa autobiografica e di "autobiografia assistita"*
13. Alessandro Bertirotti, *L'uomo, il suono e la musica*
14. Maria Antonietta Rovida, *Palazzi senesi tra '600 e '700. Modelli abitativi e architettura tra tradizione e innovazione*
15. Simone Guercini, Roberto Piovan, *Schemi di negoziato e tecniche di comunicazione per il tessile e abbigliamento*
16. Antonio Calvani, *Technological innovation and change in the university. Moving towards the Virtual University*
17. Paolo Emilio Pecorella, *Tell Barri/Kahat: la campagna del 2000. Relazione preliminare*
18. Marta Chevanne, *Appunti di Patologia Generale. Corso di laurea in Tecniche di Radiologia Medica per Immagini e Radioterapia*
19. Paolo Ventura, *Città e stazione ferroviaria*
20. Nicola Spinosi, *Critica sociale e individuazione*
21. Roberto Ventura (a cura di), *Dalla misurazione dei servizi alla customer satisfaction*
22. Dimitra Babalis (a cura di), *Ecological Design for an Effective Urban Regeneration*
23. Massimo Papini, Debora Tringali (a cura di), *Il pupazzo di garza. L'esperienza della malattia potenzialmente mortale nei bambini e negli adolescenti*
24. Manlio Marchetta, *La progettazione della città portuale. Sperimentazioni didattiche per una nuova Livorno*
25. Fabrizio F.V. Arrigoni, *Note su progetto e metropoli*
26. Leonardo Casini, Enrico Marone, Silvio Menghini, *OCM seminativi: tendenze evolutive e assetto territoriale*
27. Pecorella Paolo Emilio, Raffaella Pierobon Benoit, *Tell Barri/Kahat: la campagna del 2001. Relazione preliminare*
28. Nicola Spinosi, *Wir Kinder. La questione del potere delle relazione adulti/bambini*
29. Stefano Cordero di Montezemolo, *I profili finanziari delle società vinicole*
30. Luca Bagnoli, Maurizio Catalano, *Il bilancio sociale degli enti non profit: esperienze toscane*
31. Elena Rotelli, *Il capitolo della cattedrale di Firenze dalle origini al XV secolo*

32. Leonardo Trisciuzzi, *Barbara Sandrucci, Tamara Zappaterra, Il recupero del sé attraverso l'autobiografia*
33. Nicola Spinosi, *Invito alla psicologia sociale*
34. Raffaele Moschillo, *Laboratorio di disegno. Esercitazioni guidate al disegno di arredo*
35. Niccolò Bellanca, *Le emergenze umanitarie complesse. Un'introduzione*
36. Giovanni Allegretti, *Porto Alegre una biografia territoriale. Ricercando la qualità urbana a partire dal patrimonio sociale*
37. Riccardo Passeri, Leonardo Quagliotti, Christian Simoni, *Procedure concorsuali e governo dell'impresa artigiana in Toscana*
38. Nicola Spinosi, *Un soffitto viola. Psicoterapia, formazione, autobiografia*
39. Tommaso Urso, *Una biblioteca in divenire. La biblioteca della Facoltà di Lettere dalla penna all'elaboratore. Seconda edizione rivista e accresciuta*
40. Paolo Emilio Pecorella, Raffaella Pierobon Benoit, *Tell Barri/Kahat: la campagna del 2002. Relazione preliminare*
41. Antonio Pellicanò, *Da Galileo Galilei a Cosimo Noferi: verso una nuova scienza. Un inedito trattato galileiano di architettura nella Firenze del 1650*
42. Aldo Burresi (a cura di), *Il marketing della moda. Temi emergenti nel tessile-abbigliamento*
43. Curzio Cipriani, *Appunti di museologia naturalistica*
44. Fabrizio F.V. Arrigoni, *Incipit. Esercizi di composizione architettonica*
45. Roberta Gentile, Stefano Mancuso, Silvia Martelli, Simona Rizzitelli, *Il Giardino di Villa Corsini a Mezzomonte. Descrizione dello stato di fatto e proposta di restauro conservativo*
46. Arnaldo Nesti, Alba Scarpellini (a cura di), *Mondo democristiano, mondo cattolico nel secondo Novecento italiano*
47. Stefano Alessandri, *Sintesi e discussioni su temi di chimica generale*
48. Gianni Galeota (a cura di), *Traslocare, riaggregare, rifondare. Il caso della Biblioteca di Scienze Sociali dell'Università di Firenze*
49. Gianni Cavallina, *Nuove città antichi segni. Tre esperienze didattiche*
50. Bruno Zanoni, *Tecnologia alimentare 1. La classe delle operazioni unitarie di disidratazione per la conservazione dei prodotti alimentari*
51. Gianfranco Martiello, *La tutela penale del capitale sociale nelle società per azioni*
52. Salvatore Cingari (a cura di), *Cultura democratica e istituzioni rappresentative. Due esempi a confronto: Italia e Romania*
53. Laura Leonardi (a cura di), *Il distretto delle donne*
54. Cristina Delogu (a cura di), *Tecnologia per il web learning. Realtà e scenari*
55. Luca Bagnoli (a cura di), *La lettura dei bilanci delle Organizzazioni di Volontariato toscane nel biennio 2004-2005*
56. Lorenzo Grifone Baglioni (a cura di), *Una generazione che cambia. Civismo, solidarietà e nuove incertezze dei giovani della provincia di Firenze*
57. Monica Bolognesi, Laura Donati, Gabriella Granatiero, *Acque e territorio. Progetti e regole per la qualità dell'abitare*
58. Carlo Natali, Daniela Poli (a cura di), *Città e territori da vivere oggi e domani. Il contributo scientifico delle tesi di laurea*
59. Riccardo Passeri, *Valutazioni imprenditoriali per la successione nell'impresa familiare*
60. Brunetto Chiarelli, Alberto Simonetta, *Storia dei musei naturalistici fiorentini*
61. Gianfranco Bettin Lattes, Marco Bontempi (a cura di), *Generazione Erasmus? L'identità europea tra vissuto e istituzioni*
62. Paolo Emilio Pecorella, Raffaella Pierobon Benoit, *Tell Barri / Kahat. La campagna del 2003*
63. Fabrizio F.V. Arrigoni, *Il cervello delle passioni. Dieci tesi di Adolfo Natalini*
64. Saverio Pisaniello, *Esistenza minima. Stanze, spazî della mente, reliquiario*
65. Maria Antonietta Rovida (a cura di), *Fonti per la storia dell'architettura, della città, del territorio*

66. Ornella De Zordo, *Saggi di anglistica e americanistica. Temi e prospettive di ricerca*
67. Chiara Favilli, Maria Paola Monaco, *Materiali per lo studio del diritto antidiscriminatorio*
68. Paolo Emilio Pecorella, Raffaella Pierobon Benoit, *Tell Barri / Kahat. La campagna del 2004*
69. Emanuela Caldognetto Magno, Federica Cavicchio, *Aspetti emotivi e relazionali nell'e-learning*
70. Marco Masseti, *Uomini e (non solo) topi* (2ª edizione)
71. Giovanni Nerli, Marco Pierini, *Costruzione di macchine*
72. Lorenzo Viviani, *L'Europa dei partiti. Per una sociologia dei partiti politici nel processo di integrazione europea*
73 Teresa Crespellani, *Terremoto e ricerca. Un percorso scientifico condiviso per la caratterizzazione del comportamento sismico di alcuni depositi italiani*
74 Fabrizio F.V. Arrigoni, *Cava. Architettura in "ars marmoris"*
75. Ernesto Tavoletti, *Higher Education and Local Economic Development*
76. Carmelo Calabrò, *Liberalismo, democrazia, socialismo. L'itinerario di Carlo Rosselli (1917-1930)*
77. Luca Bagnoli, Massimo Cini (a cura di), *La cooperazione sociale nell'area metropolitana fiorentina. Una lettura dei bilanci d'esercizio delle cooperative sociali di Firenze, Pistoia e Prato nel quadriennio 2004-2007*
78. Lamberto Ippolito, *La villa del Novecento*
79. Cosimo Di Bari, *A passo di critica. Il modello di* Media Education *nell'opera di Umberto Eco*
80. Leonardo Chiesi (a cura di), *Identità sociale e territorio. Il Montalbano*
81. Piero Degl'Innocenti, *Cinquant'anni, cento chiese. L'edilizia di culto nelle diocesi di Firenze, Prato e Fiesole (1946-2000)*
82. Giancarlo Paba, Anna Lisa Pecoriello, Camilla Perrone, Francesca Rispoli, *Partecipazione in Toscana: interpretazioni e racconti*
83. Alberto Magnaghi, Sara Giacomozzi (a cura di), *Un fiume per il territorio. Indirizzi progettuali per il parco fluviale del Valdarno empolese*
84. Dino Costantini (a cura di), *Multiculturalismo alla francese?*
85. Alessandro Viviani (a cura di), *Firms and System Competitiveness in Italy*
86. Paolo Fabiani, *The Philosophy of the Imagination in Vico and Malebranche*
87. Carmelo Calabrò, *Liberalismo, democrazia, socialismo. L'itinerario di Carlo Rosselli*
88. David Fanfani (a cura di), *Pianificare tra città e campagna. Scenari, attori e progetti di nuova ruralità per il territorio di Prato*
89. Massimo Papini (a cura di), *L'ultima cura. I vissuti degli operatori in due reparti di oncologia pediatrica*
90. Raffaella Cerica, *Cultura Organizzativa e Performance economico-finanziarie*
91. Alessandra Lorini, Duccio Basosi (a cura di), *Cuba in the World, the World in Cuba*
92. Marco Goldoni, *La dottrina costituzionale di E.J. Sieyès*
93. Francesca Di Donato, *La scienza e la rete. L'uso pubblico della ragione nell'età del Web*